MINING MICROBLOGS FOR CULTURE-AWARENESS IN WEB ADAPTATION

Volume 1 of 2

Elena Alexandrovna Daehnhardt

Submitted for the degree of
Doctor of Philosophy

Heriot-Watt University
School of Mathematical and Computer Sciences
Department of Computer Science

August 2018

Abstract

Prior studies in sociology and human-computer interaction indicate that persons from different countries and cultural origins tend to have their preferences in real-life communication and the usage of web and social media applications. With Twitter data, statistical and machine learning tools, this study advances our understanding of microblogging in respect of cultural differences and demonstrates possible solutions of inferring and exploiting cultural origins for building adaptive web applications. Our findings reveal statistically significant differences in Twitter feature usage in respect of geographic locations of users. These differences in microblogger behaviour and user language defined in user profiles enabled us to infer user country origins with an accuracy of more than 90%. Other user origin predictive solutions we proposed do not require other data sources and human involvement for training the models, enabling the high accuracy of user country inference when exploiting information extracted from a user followers' network, or with data derived from Twitter profiles. With origin predictive models, we analysed communication and privacy preferences and built a culture-aware recommender system. Our analysis of friend responses shows that Twitter users tend to communicate mostly within their cultural regions. Usage of privacy settings showed that privacy perceptions differ across cultures. Finally, we created and evaluated movie recommendation strategies considering user cultural groups, and addressed a cold-start scenario with a new user. We believe that the findings discussed give insights into the sociological and web research, in particular on cultural differences in online communication.

Preface

User participation is the cornerstone of the Web today. Social networking sites facilitate the creation of real-life content reflecting user life experiences around the globe. Microblogs, as a useful tool of communication, provide an outlook into people's lives, needs, and inspirations. Knowledge about web users and their traits is invaluable for building responsive, intelligent web systems contributing to a more interesting, free for everyone and open, World Wide Web.

My work and life experience of living in a fast-paced digital world has changed drastically with the development and constant evolution of the Internet. The Network is a very intriguing and fascinating subject to study and work. The continuous demand for new applications, discoveries and exciting new possibilities to learn new things offer excellent sources of inspiration for many IT professionals, researchers, and Web users.

The primary focus of the work lies in the area of social networking sites, particularly microblogs. Microblogs are a tremendous source of user-generated content. To what extent can microblogs inform us about user context, which reflects user personality, cultural origins, and related preferences? These questions inspired me to investigate how cultural cues could be derived from microblogs, and how could we further exploit this information for building state-of-the-art applications tailored to user cultural needs.

The thesis comprises ten chapters, which can be read independently. However, for a more in-depth understanding of the main findings and the overall research quest, I would suggest starting with the first two parts which set the scene by providing the motivation and background work for building up this research. My contributions in the third part resulted in several publications, which would not have been possible without the great comments and hard work of my dear co-authors and anonymous peer reviewers' advice helping me in bringing my ideas forward.

I am very grateful for your interest in my research. I wish you happy reading!

Acknowledgements

My thesis journey took a few years filled with inspiration but challenging due to unexpected life situations out of my control. These situations helped me to learn more about academia and myself but also gain extraordinary life experience and become a stronger person, which would have been impossible without the help of great people assisting me in this life journey.

Nick Taylor was the best PhD supervisor anyone could hope. I am very fortunate to get his professional guidance and trust to develop my ideas while mastering the necessary skills amiably and intuitively. I am very grateful to Yanguo Jing for being with me these happy research years, with his expertise, cheerful support and always swift response when I most needed it. Peter kindly inspired me to go with this project and complete it despite everything opposing it. I was challenged by many anonymous reviewers, who helped me to make my publications better and stay as objective as possible with my ideas while providing much needed external feedback while researching a multi-disciplinary field. I am also very thankful to the very supportive Research Gate community, Microsoft Azure team, fellow students and conference organisers for their professional support.

I am very fortunate to have the loving and gentle support of my dear husband. Thank you, Andreas, for making things possible, so much needed support during my sleepless nights of working on paper submissions and help beyond expectations. I am grateful to my very loving and supportive family and dear friends believing in me and being with me in the hardest times possible. Thank you, my dear papa and mama, for all your love. I am very thankful to Barbara. Barbara, you are a lifesaver!

Remarkably, my research was also created with the contribution of many Web users, to whom I am indebted and with whom I made discoveries on the way. I have learned about microblogging, not only from my point of view but also with the help of thousands of users helping me in the research by sharing their content.

ACADEMIC REGISTRY
Research Thesis Submission

Name:	Elena Alexandrovna Daehnhardt		
School:	School of Mathematical and Computer Sciences		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought:	Doctor of Philosophy

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6) I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate:		Date:	30 August 2018

Submission

Submitted By *(name in capitals)*:	ELENA ALEXANDROVNA DAEHNHARDT
Signature of Individual Submitting:	
Date Submitted:	30 August 2018

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals)*:			
Method of Submission *(Handed in to SSC; posted through internal/external mail)*:			
E-thesis Submitted (***mandatory for final theses***)			
Signature:		Date:	

Contents

List of Tables

List of Figures

Glossary

API "is an Application Programming Interface that allows an application program that is written in a high-level language to use specific data or functions of the operating system or another program" as defined in [118]. For instance, Twitter API allows access to and search of Twitter resources.

CF Collaborative Filtering is one of the most popular techniques for creating recommendation systems. As defined in [118], CF is a "personalization technology that calculates the similarity between users based on the behaviors of a number of other people and uses that information to make recommendations for the current user."

DT Decision Trees is widely used machine learning technique for addressing supervised classification problems.

GPS Global Positioning System, is a global satellite system providing an information on geographic coordinates.

ICT Information and Communications Technology.

IMDB The Internet Movie Database is the online information source of movies, television and other entertainment shows, which can be reviewed and rated by individual users.

LLM Latent Language Models, are the statistical models learning hidden (or latent) words based on a corpus of text documents.

locality We define locality to any location-dependent data we retrieve from the Twitter. We mostly refer to user time-zone, free-text location field, geographic coordinates and even preferred language as locality traits, with which we could further infer tweet origins.

Metadata is data that narrates other data. We refer mostly to Twitter tags used for structuring and describing tweets. For instance, user mentions include screen_name reflecting an online alias of the user.

microblog "A service for posting short, public messages. A microblog message is usually one or two sentences long. It is useful for many purposes such as sharing links, asking questions, and making statements." as defined in [118].

microblogging "A method of blogging that limits bloggers to a fixed number of characters. Forms of microblogging include tweets and text messages." as defined in [118].

origin We mostly refer to the tweet origins including inferred countries and regions, called also cultural dimensions in accord with R.Lewis sociological research [148]. It is important to mention that tweet origins might not match particular user country of origin or cultural dimensions associated [1]

REST REpresentational State Transfer, defined by Dr. Roy Fielding in [71].

RS Recommender Systems are software applications for suggesting their users on products and interesting content such as news and movie advertisements.

SN Social Networking, we refer here to social networking web applications including Twitter and Facebook.

TCP/IP Internet Protocols, the Transmission Control Protocol (TCP) and the Internet Protocol (IP).

TF-IDF Term Frequency-Inverse Document Frequency, an information retrieval metric for defining a word's (term's) importance to a document in a text corpus.

tweet "A post made on the social media application Twitter.." as defined in Oxford University Press [182].

URL Uniform Resource Locator, a web address pointing to a web page or other web resource.

[1] Table H.3 presents our human assessment of Twitter profiles with geographic coordinates available in tweets. As seen from Table H.4 (a), the country assigned by Twitter and Rater 1 agreed in 83.93% (Krippendorff's $\alpha = 0.81$), and by Twitter and Rater 2 agreed in 78.57% (Krippendorff's $\alpha = 0.76$). Please refer to Methodology page 74 for the inter-rater reliability coefficients' overview including Krippendorff's α.

WWW World Wide Web is the global Network of interlinked information resources and communication services.

List of symbols

MU Rating prediction approach considering an overall average rating.

$BOOSTER$ Movie rating prediction technique based on gradient boosting[2] regression model using "weak learners" or "shallow trees" while optimising the prediction error.

$FACTORS$ Rating prediction technique using factorisation machines as described in [199].

$ITEM$ Rating prediction approach considering an average rating computed on all ratings given to the movie.

$OFFSET$ A baseline rating prediction approach considering user and movie average ratings and related deviations as described in [136, 63].

$PLACE$ is a feature set used for creating user country predictive models using Twitter information on user language defined in one's profile, timezone and free-text location as an arbitrary text provided by microbloggers.

$USER$ Rating prediction approach considering average user ratings to the movies they rated.

[2]See description at scikit-learn

Publications by the Candidate

[51]. E. Daehnhardt, N. Taylor, and Y. Jing. Mining Microblogs to Exploit Culture-Awareness in Web Adaptation. *SICSA PhD Conference*, University of Glasgow, 2015.

[53]. E. Daehnhardt, N. Taylor, and Y. Jing. Usage and Consequences of Privacy Settings in Microblogs. *International Workshop on Social Media Mining and Analysis*, IEEE: 467-674, 2015.

[52]. E. Daehnhardt, Y. Jing, and N. Taylor. Cultural and Geolocation Aspects of Communication in Twitter. *Social Informatics*, ASE, Academy of Science and Engineering: 1-12, 2014.

The following publications were published under my previous name "Ilina":

[121]. E. Ilina. A User Modeling Oriented Analysis of Cultural Backgrounds in Microblogging. *Human Journal*, I(4): 166-181, 2012.

[119]. E. Ilina, F. Abel, and G.-J. Houben. Mining Twitter for Cultural Patterns. *ABIS 2012 Workshop on Personalization and Recommendation on the Web and Beyond*, editors: Reiterer, Harald and Deussen, Oliver: 83-90, 2012.

[120]. E. Ilina, C. Hauff, I. Celik, F. Abel, G.-J. Houben. Social event detection on Twitter. *Web Engineering Conference, Berlin*, Springer: 169–176, 2012.

Part I

Motivation

Chapter 1

Introduction

1.1 Background

"The beginning is the most important part of the work."

- Plato, The Republic

Social Networking (SN) applications including Twitter microblogs, LinkedIn for job search and online solicitation and Facebook for communicating with family and friends are used by millions of users worldwide. About a quarter of the World population employs SN in their daily life, communicating with real-life friends and online acquaintances [9]. In April 2018, these web applications were amongst the top 16 social networks with the active user base of more than 2 billion of users for FaceBook (leading), 330 and 260 million of users for Twitter and LinkedIn respectively [231]. Furthermore, Twitter users are very active, sharing hundreds of millions of tweets daily [173].

Consequently, user-generated content can be used to broadcast life news, explore real-life events [255], find trending topics [96] and mine user interests [275] or opinions [279]. A deeper understanding of user interests and traits, technology preferences, etc. could be used therefore for an improved user experience online, e.g., personalised web applications. For instance, information on user locations and culture-specific preferences on food could be used in urban planning [225]. An e-learning environment would benefit from the knowledge of user cultural origins since learner expectations towards study materials and teaching methods differ amongst cultures [180]. With knowledge of user cultural origins, adaptive applications can be tailored to specific culture-related user traits.

1.2 Problem Statement

Culture-aware web adaptation has potential benefits for applications in e-commerce, e-learning and SN. Users could be provided with easy-to-use web interfaces and web content designed with user cultural preferences in mind. There are findings on cultural differences in user behaviour and preferences online [225, 190, 56, 127]; however, the ways of automatically mining user cultural origins and outcomes of exploiting this information for delivery of personalised content and functionalities are not yet clear.

For instance, would cultural user traits be advantageous to consider in recommendation engines? Could we build on foundations of the existing sociological studies (such as Hofstede) while creating state-of-the-art web applications? What could be the implications of adapting to cultural behavioural differences online?

The object of this study is user interactions in online microblogs. Particularly, we focus on Twitter microblogs and seek to find out how the cultural context of Twitter users impacts user microblogging activities. The main focus is thus on the following questions:

- How do users exploit microblogs and do cultural differences play a role in microblogging activities?
- Could the differences found be exploited in web applications? For instance, would a recommendation system benefit from a knowledge of user origins?

1.3 Contributions

The main research contributions are:

- an analysis of Twitter feature usage for identifying distinctive microblogging patterns in respect to geographic locations of Twitter users;
- exploitation of these microblogging patterns to build up predictive classification models of user origin;
- analysis of microblogging communication preferences for users tweeting from the most active countries in Twitter;
- analysis of privacy controls usage in Twitter;
- exploitation of inferred user origins for the design and evaluation of culture-aware social recommendation strategies.

1.4 Research Scope

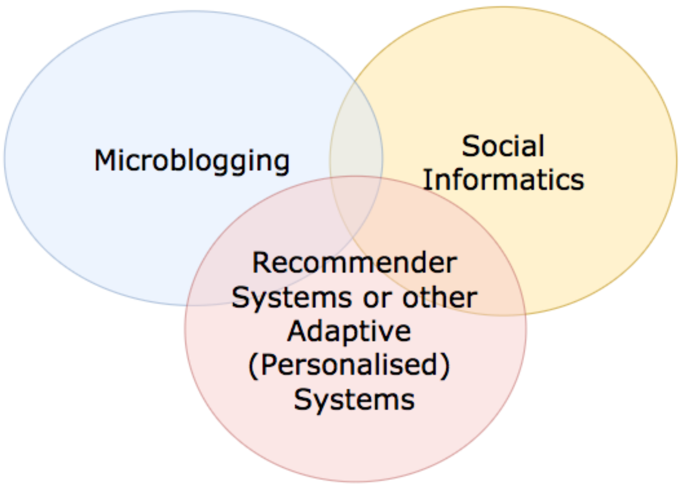


Figure 1.1: Research scope is in the conjunction of several knowledge fields

This research is interdisciplinary and lies in the intersection of recommender systems, microblogs and social informatics (Figure 1.1). Overall, it employs a socio-technical approach and sociological studies when analysing user interactions online, building on previous research foundations outlined in the second part of this thesis (chapters 2 to 4). In the following third part, the main research objectives and methodology are explained (chapter 5), followed with research contributions (chapters 6-9, based on published works [119, 121, 52, 53, 51]) and discussion of results (chapter 10).

In particular, this work addresses user behaviour analysis for implicit data collection, to address the cold-start problem in cases when user origins are required, but not known to the system. Once user origins are inferred out of user-generated content and metadata, this information can be exploited to analyse user communication preferences (chapter 7) and build culture-aware social recommendation strategies (chapter 9). In chapter 8, existing studies in the research area of Twitter microblogs mining and privacy issues in respect to using Twitter privacy controls are reviewed. The concluding part, chapter 10, critically analyses the results in respect of the previous research, points out limitations and proposes directions for further study while referring to practical applications which could benefit from the research findings.

In a nutshell, SN platforms such as Twitter provide large volumes of user-generated content, which while dealing with information overload, could be exploited in view of sociological research to better understand user behaviour online. While cultural cues might be found online, their exploitation for providing culture-tailored adaptations is not yet studied in-depth. How could we address different user needs following their culture-related user preferences? How could we mine cultural user traits from the openly available user-generated content? How could we exploit knowledge on user cultural backgrounds for creating responsive state-of-the-art web applications? In the next chapter, I start to address these questions, discussing prominent research works in the areas of social informatics, microblogging and recommender systems.

Part II

Background

Chapter 2

Social Web: Challenges and Opportunities

"The Web is a social technology that thrives on growth ..."

- Tim Berners-Lee et al. [22]

2.1 Introduction

In this chapter, we will introduce the fundamental concepts of the Social Web and Networking. Specifically, we focus on the most notable research papers on microblogging, since one of our main aims is to identify microblogging behavioural differences of specific user groups defined by their geographic locations as a proxy for their cultural origins. The existing context-aware web applications and recommendation systems are herein critically discussed regarding the existing research.

2.2 Approach

Social Informatics and Networking. As seen in Figure 2.1 below, we will start from the general principles and move on to more specific studies related to our research objectives. While selecting our literature sources, we will focus on prominent[1] research papers in the area of social informatics and networking. The

[1] in terms of the number of citations and subject related to the research direction

social informatics perspective allows us to understand how technology and society influence each other. While information systems and technological advancements were seen before as supportive tools for achieving personal and work-related goals, the Web evolved into a new platform for social interaction and networking, in which user needs for sharing information and collaborating with other users were met.

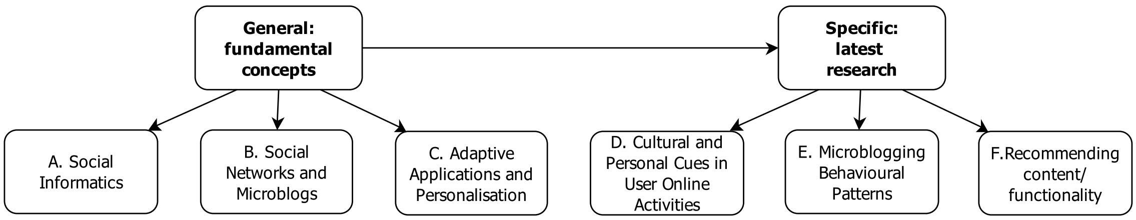


Figure 2.1: Literature review process

User Context Usage. When considering a user-centred design, state-of-the-art web applications might require knowledge on user origins and preferences. For instance, user locations and used languages could be exploited for providing related content and functionality. Further, we will concentrate on the latest developments in the field of recommendation systems, particularly in context-aware Recommender Systems (RS). Our main interest is to use openly user-generated content provided by Social Networking services such as micro-blogs for mining user data, which might be further exploited while creating user profiles. We will focus on extracting and exploiting culture-related user traits and context in view of recent research findings pointing to online behaviour heterogeneity for users coming from different nations and cultural groups. Our main question is to understand whether user cultural context could be useful for improving recommendation performance or adapting web applications to specific culture-related user needs.

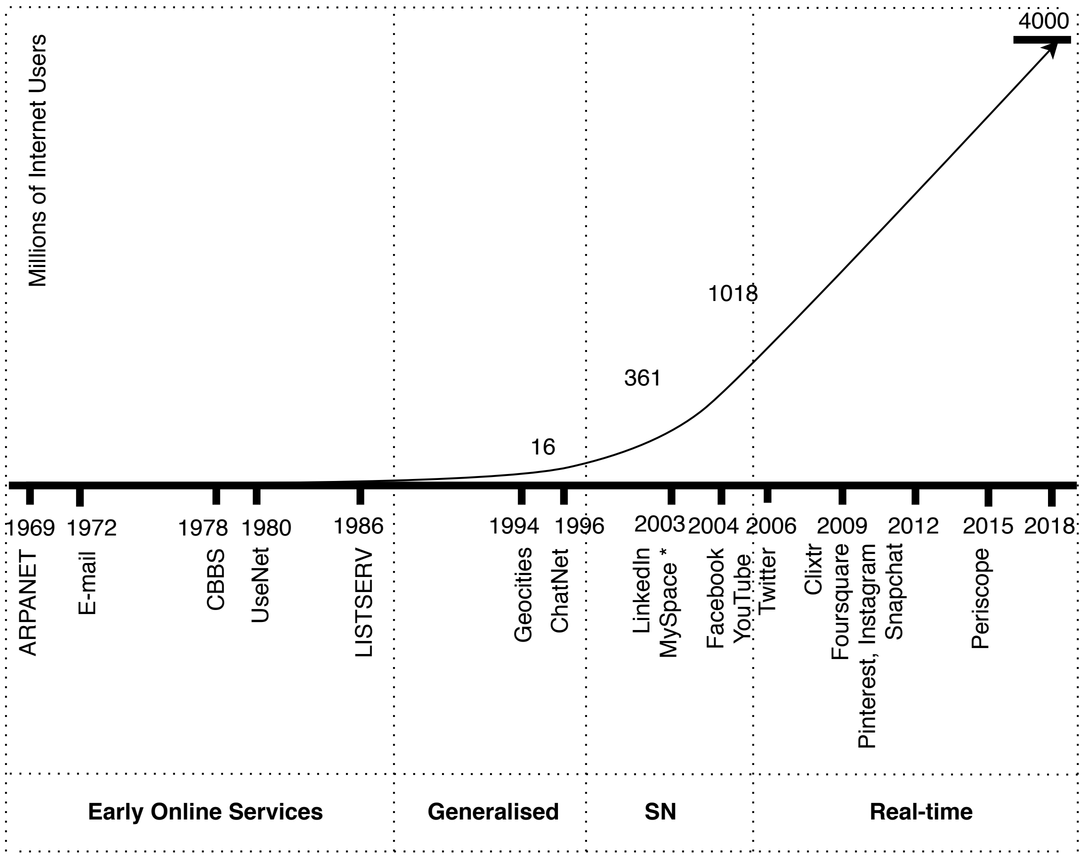


Figure 2.2: Online services evolution and growth [124] of their user base (based on the "A Brief History of the Internet" by Leiner et al. [146], the user base statistics provided at internetworldstats.com [124] and Wikipedia.org sources on the social media and networking applications [270, 268, 267, 264]

2.3 The Internet and World Wide Web

2.3.1 Evolution of The Network and Services

The Internet is a global network of networks consisting of computers and other electronic and mobile devices using the Internet Protocol suite (TCP/IP) for communicating information. The main idea of the Internet was to provide a reliable global network, allowing connectivity without regard to the underlying architecture of the connected networks [146]. The "openness" and reliability of the Network is the cornerstone of Internet development, from the early creation of ARPANET in 1969 till today. The initiative to build robust and fault-tolerant computer networks resulted in the de-centralised global system of interconnected computer networks [146]. The Internet as we see it now went through the development of various hardware devices, protocols and services enhanced over the past decades (see Figure 2.2).

This led to the wide adoption of the web services, with the rapid user base increase towards 4 billion at the end of the 2017 year in accord with the statistics provided by internetworldstats.com [124].

The ARPANET was one of the earliest packet switching networks and was developed with the support of the Advanced Research Projects Agency (ARPA, later DARPA) within the U.S. Department of Defence in 1969 [146, 263]. The ARPANET was used in US research laboratories and universities and led to the development of the global Internet network with the implementation of the Transmission Control and the Internet Protocols (TCP/IP) [146]. The development of TCP/IP enabled the first Internet applications such as electronic mail, voice transfer and sharing of computer resources [146]. The TCP/IP was subsequently integrated into operating systems, resulting in further acceptance of the network [146]. While the network was initially designed for military and research purposes, in 1985 the Internet started to become available to the general public [146], and the number of interconnected networks started to grow.

The Internet provides a technical infrastructure to support e-mail and instant messaging, file transfer, newsgroups, online games and other online services. The Internet and WWW enabled a platform for development and provision of online networking services. The early online services allowed web users to upload and download files, read and exchange news, send messages and chat with other users with the help of Computerised Bulletin Boards (1978). Discussion services such as UseNet (1980) enabled early Internet adopters to participate in online discussions using the terminal software. While the first electronic mail message was sent at the beginning of 1970s [146], in 1986 the electronic mailing list provided by LISTSERV enabled users to send messages to other web users subscribed to related mailing lists.

For user-friendly access to online services and resources, Tim Berners-Lee proposed the development of a global hypertext system to improve knowledge management at the European Organisation for Nuclear Research (CERN). The main idea was to access information sources and user-generated content via interlinked hypertext documents published online. These documents could contain textual and multimedia (images, video, podcasts, sound) content and are called web pages. For easy access to the web pages, Berners-Lee started the development of the first visual browser for the WWW in 1990 [21]. Overall, the availability of these communication protocols, technologies and web browsers resulted in the rapid growth of various web applications and services in the following years.

Further, online services were created to support generalised online communities such as Geocities (1994), The Globe and Tripod.com (1995), and ChatNet (1996). The emphasis of these applications was on sharing personal ideas and information in chat rooms or discussion boards. As more and more users went online, finding friends and creating online contact networks became pertinent for web users. This led to the development of websites such as SixDegrees.com (1997), FriendSter (2002), LinkedIn (2002) and MySpace (2003), and Facebook (2004). Web users were enabled to communicate with their real-life acquaintances, create their user profiles and form online connections.

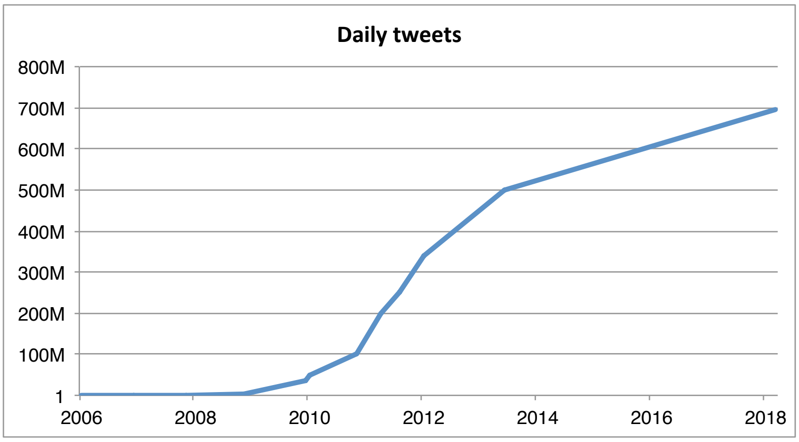


Figure 2.3: Volume of Twitter messages increased over time
(based on the data provided by internetlivestats.com [123, 122])

The wide adoption of SN websites and ease of information exchange with the help of web technologies resulted in the development of real-time broadcasting services. Generalised online sharing services developed mainly in the 1980s and 1990s grew into more user-centred Social Networks and real-time applications, such as Facebook and Twitter (in the 2000s). Users were enabled to share a video on YouTube (2005), textual information on Twitter (2006), geo-tagged photos on Clixtr (2009) and share their locations with friends on Foursquare (2009). As for April 2018, sharing visual ideas and content of other contributors in real-time was used by 200 million users with Pinterest (launched in 2010), Instagram (2010) was used by 813 million users for sharing own content, and Snapchat (2011) allowing its 255 million users to share visual messages, which can be edited with filters and become unavailable with time [231, 270, 138]. Twitter microblogs become rapidly popular, and the number of daily Twitter posts increased from a first tweet in 2006 to 500 million tweets in 2013 as stated in [123], and reached about 700 million of tweets as for May 2018 [122] and shown in Figure 2.3 above. Even though Twitter microblogging service was initially designed for communicating with textual messages (known as "tweets"), in 2015 it bough "Periscope" web application for life video streaming and commenting shared content [271, 221, 270]. It shows that Microblogging services such as Twitter are likely to evolve while adding more features and applications for multi-media user experience not limited to short messages.

What would be next developments in online services in future is difficult to investigate due to a large number of social application startups providing customised solutions for any taste or need. For instance, plane sharing platform wingly.io [49], which won Startup Battlefield Europe in 2018 [48], allows its users to make personal flight requests from the registered licensed pilots. Several prominent directions of online services development, however, emerged during the past few years. For instance, the development of cryptocurrency facilitating the fast adoption of social media platforms that reward their contributors for creating compelling content, such as proposed by Mithril Taiwan-based startup [167]. Another possible direction could be an adoption of Artificial Intelligence (AI) techniques for creating more "user-friendly" social tools such as chat bots, for instance, the most recent application at "huggingface.co" using language models using meta learning and having different memory scales [274]. The modern chatbots, however, require more adaptation to address ethically-challenging real-life situations. For instance, a recent work [50] revealed different types of chatbots' responses to sexually insulting user requests, resulting in void, inconsistent and even provocative system replies. AI tools and big data exploitation is more likely to happen in near future when we think about the recent Facebook acquisition of the Face Recognition tool [98]. To analyse the big data, big data processing solutions recently become attractive for acquisition by big Internet companies such Google as we see from the "Cask" startup enabling a seamless usage of Hadoop and Apache Spark tools [47]. However, the privacy ramifications cannot be overlooked due to the recent issues of user data leakage as discussed in [46], which can be seen as an example of how social media could be abused on the larger scale when the user data is not safeguarded properly. The need of better user control is recently brought to focus in European Union for privacy [2] and data protection regulations to bring more transparency over shared user data [65, 66].

2.3.2 Social Informatics Perspective

With a large number of users and user-generated content, SN provides an attractive platform for research in the fields of interpersonal online interaction, computer-mediated communication, and social informatics. SN services are complex systems which can be seen from a socio-technical point of view combining the technical aspects of SN with human psychological and social factors which play an essential role in personal communication online.

Moreover, social networking could be further enhanced when human social interactions online are supported with pervasive computing techniques such as information from neighbouring devices. This way we could build on advantages of both technologies as discussed in [186]. A developed a pervasive social networking environment experimental setup allowed to analyse user attitudes in user trials, resulted in overall positive user feedback [186].

One of the notable definitions of Social Informatics (SI) as an interdisciplinary field was provided by Kling [134]. SI investigates the design and application of information and communication technology while considering its reciprocal influence on social groups within particular cultural settings. The various questions arising

[2] We analyse Twitter privacy settings usage in chapter 8 "Privacy Settings Usage in Twitter"

from this mutual influence are from different research directions within SI; how technology influences society and how society affects the technology. Kling [134] looks critically at the deterministic research questions first asked in the 70s-80s when performing organisational research and studying the impact of ICT on organisational behaviour. To avoid "oversimplification", Kling [134] suggests studying ICT within its related sociological context as a "sociotechnical network". He [134] argues that technological tools and affected societies are intertwined and influence each other.

For instance, Internet Technologies (IT) are interlinked with the society using it which influences its design and use. The social context of IT applications influences the social ways in which persons use IT, while society creates the IT infrastructure, develops required skills and establishes IT's social usage context. Kling [134] refers to sociological, cultural, organisational and other aspects playing an inseparable role in ICT advancements. The main goal of any information system is to support the work and life of its users [134].

2.3.3 Social Interactions on the Web

User-Centered Networking. While various Social Networks such as Facebook, MySpace [3] and LinkedIn have become increasingly popular, the Internet environment provides a richer means for user communication and information exchange. The Social Web is defined as a collection of web applications and services in which user participation is a cornerstone [93]. Web users are seen as creators and final consumers of the information including textual and multimedia sources. The Social Web is the place where people share information with each other and learn from each other. According to [93], the "culture of mass participation" is inherited into Social Web sites. Moreover, SN usage and development is a hot topic in information systems research. For instance, SN usage by students is explored in [202], expression of political opinions online is discussed in [279], SN and life connections on MySpace, Facebook, LinkedIn, Twitter are studied in [99].

It is important to mention that even though these social networking tools al-

[3] MySpace was "re-launched" in 2013, and with its parent company Viant, was bought by Time Inc in 2016 [270, 230]. In [270]

low their users to form friendship connections, they are not always bi-directional or reciprocal as discussed in [144]. When Twitter requires friendship reciprocity, Twitter connections are not always reciprocal. In the finding of Kwak et al. [144] only 22% of users connected in ties, which is much lower than for other services such as Flickr having 68% rate of reciprocity (referring to the previous work [36]). Kwak et al. [144] explain it by the particular nature of Twitter, which used mostly as an "information sharing" tool.

Social Media Usage. Kaplan and Haenlein [130] defined Social Media (SM) as a group of Internet-based tools enabling the sharing and transmission of user-generated content. They [130] emphasised the importance of SM for businesses willing to exploit the SM benefits and provided advice on how to select and use the SM respectively for business needs and communication effectiveness. While the communication effectiveness is higher in face-to-face communication, and lower in weblogs, web users might select their preferable type of SM not only to communicate effectively but also regulate the level of personal information sharing. For instance, the level of personal sharing is lower for collaborative projects, while SN sites enable higher personal information sharing for their users [130]. To summarise, SM can be seen as a set of enabling technologies for information exchange, while SN sites such as Facebook and Twitter can be referred to as the process of using SM. A Social Networking Service was defined in [127] as "a web-based service that forms relationships between individuals by providing profile bases which include individual information, makes social interaction between participants easy by providing users with functions for communicating with each other, and provides a platform for users to share information and contents."

2.3.4 State-of-the-Art Web Applications

With the growth of the number of users and user-generated content, the vast amount of openly available information becomes difficult to navigate and further exploit without additional developments in the fields of web search [27], data-mining [44, 137, 62], machine learning [217] and recommendation system [152, 168]. Some of these techniques allow us to filter out or hide irrelevant information or functionality while providing personalised experiences benefiting web users and e-commerce.

Adaptive Applications and User Preferences. State-of-the-art web applications such as e-learning environments require knowledge about their users, for instance, students' study goals, acquired competence and current interests, and can gather relevant user traits for adapting or personalising user interfaces for delivering relevant content or functionalities. For example, an e-learning system can show only courses to which the student has prerequisite knowledge. To provide a better user experience, adaptive applications require additional information on user traits, interests, goals, competency and other individual characteristics which can be stored for further access to enable adaptive behaviour [29].

User Models and Modeling. User traits can be organised in user models represented by user profiles within the realised system [29]. The user characteristics can be obtained directly by asking the user or gathered implicitly by analysing user activities within the system. System logs, usage patterns, navigation activities and purchase history can provide an outlook into user characteristics and be further used for building up user profiles [62]. To be more precise,Ffischer [72] defines user models "as models that systems have of users that reside inside a computational environment". The information stored in user models is thus application-specific and contains characteristics related to the usage domain.

The process of collecting and processing user data is called "User Modeling" [29]. In [128] we read that "the goal of user modeling may be to predict user behaviour, to gain knowledge of a particular user in order to tailor interactions to that user or to create a database of users that can be accessed by others. The goal of user modeling may even be the creation of the model itself when that model is used to create an autonomous agent to fill a role within a system." Overall, the user modeling field deals with scientific and practical issues concerning gathering and maintenance of user profiles. User profiles can be exploited in various web systems requiring adaptive behaviour, for instance, when providing relevant content, hiding irrelevant information or assisting in system navigation [28].

The user modeling process involves user data collection, assessment or inference of user traits and further adaptation or personalisation implying the usage of the model [29]. User models can be created based on a domain model, which is filled out for each user. A domain model reflects a real-life world model and its concepts, which are used in an adaptive system. Alternatively to analysing activities and preferences of particular users, a group-based approach can be employed when adaptation outcomes are derived from the stereotypical user groups with which the system associates its users [135]. Stereotype-based models provide a very simple user categorisation into user groups called stereotypes. In accordance with the stereotype assigned to the user, a user is provided with related functionality and/or content. Overlay user models are more complicated and can provide better personalisation compared to stereotype-based models. To tackle uncertainty when inferring user characteristics, approximate reasoning techniques can be exploited for creating uncertainty-based user models. These modeling techniques can be combined to store several user stereotype parameters which can be handled with uncertainty in mind [29].

The most recent development in human-computer interaction field is "pervasive" or "ubiquitous" user modeling, defined in [105] as "ongoing modeling and exploitation of user behaviour with a variety of systems that share their user models". In his thesis "Ubiquitous user modeling" [105] Dominikus Heckmann discusses privacy and machine processing with the help of semantic ontology creation issues of pervasive user modeling. Due to various heterogeneous information sources of user characteristics, including also sensor data on user context and location, and needs for secure and portable information exchange, on their World Wide Web Consortium web page "User modeling" [169] Mohamad and Kouroupetroglou emphasise further standardisation of the user modeling ontology. They also re-defined user models in a broader meaning as "explicit representations of the properties of an individual user including needs, preferences as well as physical, cognitive and behavioural characteristics" [169]. Therefore, the user models can be finely tailored to personal user requirements including behavioural and mental concepts.

2.4 Mining User Preferences and Context

2.4.1 User Preferences and Context Elicitation

Explicit or Implicit Context Provision. Following the system design concepts, user characteristics and environment can be provided by users explicitly or implicitly [113, 7]. When the user-related data is indirectly (implicitly) extracted out of the system's history logs, transactions and activities or user-generated context (for instance, user opinions and rankings), then different data mining and machine learning techniques can be employed. For instance, statistical language and classification models can be applied to review blogs (such as Epinions.com) for inferring user opinions and sentiments towards reviewed products [183].

Cold-start problem and Social Web. In the case when user information is not available, either from an explicit request to the user or their usage logs, as an alternative for dealing with this "cold-start" problem, social web data can be used to gather user preferences [280]. User interests derived from microblogs such as Twitter can be used for creating user recommendation systems [2]. Twitter streams can provide more up-to-date content when required to learn about user interests [1]. However, microblog mining is a relatively new research area, with its opportunities and challenges as discussed in [151]. In the next chapter, we will focus on the data mining of Twitter microblogs and present an outline of the widely exploited methods such as referred in [151, 277].

2.4.2 Geographic Locations vs. Cultural Origins

Location Detection. Modern adaptive applications can also require information on user environment, technical details of web browser and hardware parameters such as resolution and a more finely defined social context of a user [29]. In this case, user features can be perceived as "long-term" and the immediate environment of the user as "short-term" features stored as a set of names and their values [29]. One of the examples of the context information is users' location data, which might be required for mobile systems recommending content and activities specifically personalised for the user's geographical location.

User location information can be used to provide location-specific content [194] and applications tailored to specific user interests [39]. When available, Metadata on user geographic locations can also be easily retrieved from publicly open user profiles. Explicit user location, however, is often missed or not accurate [104], and only a minority of openly-available micro posts are geographically tagged [104]. This is why location detection within the social web is a pertinent research topic.

One method to detect locations from micro post (or tweet) content, is the usage of a gazetteer or toponym vocabulary comprised of location-specific terms. However, the application of gazetteers for location information disambiguation in microblogs is challenging due to similarly named locations and inherent difficulties of mining information out of microblogs. Misspelling and usage of abbreviations are common due to the short message limitations [278]. To improve the performance of location detection using the GeoNames gazetteer, [278] employ a Support Vector Machine classifier using Twitter features extracted from tweet content (toponym mentions) and metadata (extracted from the Twitter place fields).

As assessed by human annotators, geocoding services Yahoo and Google applied to profile locations on Twitter are not reliable for a large proportion of tweets [90]. This is unsurprising since about 30% of users do not provide an accurate location [104]. As [100] pointed out, user location is influenced by temporal dynamics and requires a model update when used as a feature in a location detecting classifier. Named-entity recognition with Stanford NER and Open NLP tools is investigated in [153], showing a considerable performance when training on Twitter-specific content, which requires human involvement for annotating data. Location disambiguation of Tweets with Stanford NER, gazetteer, heuristic rules was performed with precision and recall of around 80%, and is comparable with human annotations [85]. Authors also suggest that representation of the tweets' content with the help of ontologies

[85] might assist in the toponym disambiguation task.

The detection of the home country from Twitter content was also investigated in [104] with a machine learning technique, whilst analysing tweets' content and disregarding other Twitter features and the extended contacts' network of the users. Using a Naïve Bayes classification model working with term frequencies, user country locations were inferred in about 73% of cases [104]. Geographical topic models based on terms extracted from content correlate with specific geographic locations, but require an adjusted probability estimation with the help of a smoothing technique to deal with term sparsity [203, 39, 40]. Location-specific term selection with Kernel Density Estimation is investigated in [252], demonstrating the robustness of the approach of geotagging Flickr photos when only a small set of terms is employed.

In [100] authors exploit location, username, description and time zone fields for creation of a location-detecting classifier, finding a country location with an accuracy of 92% for the best feature set analysed. They analysed the generalisability of training on a set of users sharing their geographic locations while observing a positive outcome in tests considering geo-tagged and not geo-tagged training datasets [100]. Mahmood with co-authors [156] created an ensemble classifier for detecting users' home location based on words and hashtags extracted from tweets, tweets' frequency dynamics and a gazetteer dictionary of geographic place names. Their classification algorithm enabled hierarchical location detection of time zone, state and city name with recall figures of 0.78, 0.66 and 0.58 respectively. For improving location detection outcomes for Flickr images, [103] use statistical inference, gazetteer and other features extracted from the Flickr users' metadata and content.

In location disambiguation based on user-generated content, adding content of social friends helps to improve prediction performance [125]. Locations of users sharing web links help to predict the link origins [43]. Locations of users can be predicted based on their contact networks, also considering several social platforms [129]. Zheng with co-authors [282] provided an overview of the previous research works on geographic location prediction on Twitter and suggested usage of deep learning and neural networks techniques as possible future research directions.

Cultural Backgrounds. However, when creating adaptable interfaces, location-based profiles might not be appropriate for individual users. For instance, Google's location specific redirection might not be suitable for a user whose origin is one country but who is currently residing in another country. In more complicated cases, for instance in distance learning systems developed for students coming from different cultural backgrounds, it could be important to acquire knowledge on the users' origins [119]. Such systems require not only knowledge on current location, but also an understanding of cultural origins of the persons. As a possible culture-related metric, Reinecke and Bernstein [196] proposed to employ knowledge on time spent in a particular country. Other traits such as travelling, living abroad or education could contribute to culture-awareness of users [198].

What is culture? Reinecke, Schenkel, and Bernstein in their article on user modeling [198] have pointed out about some challenges in defining a terminology for culture. They stated that in light of current globalisation and migration processes, it is difficult to define the term culture. Employing cultural traits based on country locations for personalisation of software applications might thwart the personalisation efforts due to an inherent complexity in the meaning of culture.

Sociological Models of Cultures. Applied to Information Systems and Informatics research, sociological theories are explaining cultural behavioural differences usually based on surveys of countries and employing cultural dimensions to stereotype individual behaviours. These theories include Hofstede's Cultural Dimensions [111] and the Lewis Model of Cultures [148] [4]. Even though the limitations of creating national stereotypes and overall sociological model's usage are critically discussed [163, 273], their application is considered, for instance, in e-learning [190], e-commerce [56] and social networking [127] research.

[4] both models are dsicussed in chapter 4

2.4.3 Cultural Behaviour Patterns Online

Even though existing approaches can be used for gathering information on particular user traits, only a few selected works investigate how user microblogging behaviour differs between distinctive cultural user groups. Caution is required when assuming homogeneity in user behaviour across cultures when referring to cultural models based on strong and possibly biased assumptions, as suggested in [163]. An effect of organisational culture with possibly non-uniform characteristics, the individual variations of personalities within the analysed cultural groups should not be underestimated when performing generalisations based on the survey results within an international organisation [163]. However, recent research findings demonstrate differences in online behaviour for people coming from different cultural origins.

A comparative analysis of Japanese and English blogs was performed by Nakasaki with co-authors [172], identifying opinion differences between both cultural groups on selected topics, using Wikipedia as the reference. Trending topics analysis was performed by [272], revealing international differences of users interests in news and topic popularity across cultures of six countries based on a large dataset of tweets. Tweets as a source of information on music genres popularity were studied by Schedl and Hauger [211], finding that user preferences differ amongst countries and cities. On the Facebook social networking website, functionality employed and usage time differ across cultures [254].

Observing Twitter connections within national and international geographical boundaries, Takhteyev with co-authors [237] discuss the nations as defining cultural representations of communities, in which people communicate differently due to their different interests and geopolitical status of their countries. There are patterns of follower-friend relationships, in which US users are usually followed by users from other countries, while Japanese are usually followed mostly by Japanese [237]. Interestingly, the majority of connections are within national borders [142, 237]. In accord with [237], 39% of connections are within 100 km distance. English-speaking users also follow English-speaking users in the majority of over 90%, while other language users build connections with users of the same language, in 60% of cases

[237]. Poblete with co-authors [193] studied microblogging behaviour by analysing language usage and network-related features for ten countries on Twitter. They [193] stated that the USA is the first country in their list, leading in URL sharing.

2.5 Exploiting Context Awareness

2.5.1 What is Context?

When creating web applications better tailored to specific user needs, it can be useful to consider user circumstances [7]. User environment traits can include weather observations, place and access time amongst other things. There are various meanings of context as discussed in [7] referring to definitions proposed in [61], which can be seen as static and independent from user activities and interactions with the system, and dynamic context, which changes when influenced by user activities.

Bielikova with co-authors in [25] stated that user context information such as geographic location can also be exploited in the recommendation process. When user location data is provided by mobile devices with the help of web browsing agents, the user's immediate location can be used in the user modeling process. This way, however, we have short-term knowledge on user location, and we might not know if a user is located in the country of origin or is travelling. When long-term user profiling is required, past user locations can be extracted from user activities on the social web.

Web applications such as eventful.com or upcoming.yahoo.com can shed light on user activities in real life and thus their geographic locations. In [120] we proposed an approach to social event detection from Twitter microblogs. The automatic filtering of microblogging content can be used for gathering related content and further enriching user profiles with real-life events which are interesting for the user.

When recommender systems are designed to include other parameters influencing user ratings of the items they are "context-aware" [7]. The context information can be explicitly provided by users or inferred from their activities [7]. For instance, users' geographic location context could be exploited for providing language learners with places of interest, in which learners could communicate with native speakers [82]. In Twitter, users' country-wise similarity could be used for generating friendship recommendations based on suggestions by the network [83].

Besides user location context, there are research works investigating adaptation and personalisation outcomes when also considering the cultural backgrounds of users. In e-learning, [180] found that the cultural context of a learner might impose requirements on technology usage, selection of media and the style of interaction between students and instructors. The learning material presentation methods impact the performance of students from different cultural groups [262]. To improve the experience of students in e-learning environments, it is thus paramount to distinguish between different cultural preferences [180].

In a recommendation system context, Silva with co-authors [225] analysed recommendation prospects for urban planning in accordance with user cultural similarities on Foursquare and user preferences. Cultural behaviour differences in user activities on Twitter were investigated in [80], suggesting to exploit such differences for building state-of-the-art online communication tools and friend recommenders. The usefulness of context-aware restaurant recommendation was studied in [253] with the help of user surveys. Even though the majority of participants were positive about the usefulness of the recommendations, they indicated a desire for better control of the decision making aspect in the recommendation [253].

2.5.2 Including Context into Recommender Systems

In order to include contextual information into Recommender Systems (RS), [7] suggested the following approaches:

- Contextual Pre-filtering, which works similarly to traditional recommender systems, and rates the pre-filtered items in accordance with predefined contextual variables.

- Contextual Post-filtering, in which recommendation results are filtered out based on contextual requirements.

- Contextual Modeling, in which ratings are calculated with the contextual information factored into the models.

The strategy of employing post-filtering and pre-filtering approaches is discussed in [185]. Karatzoglou with co-authors [131] propose a tensor-based model to efficiently represent the contextual data while achieving a smaller mean absolute error in recommendations compared with recommendation approaches which do not consider context.

2.5.3 Recommendation Techniques' Evaluation

Recommendation Task. RS are software applications designed to infer user preferences towards specific items and their properties. For instance, RS can help in recommending movies (MovieLenz), books (Amazon), music tracks (Last.fm) and other products or services. There are various recommendation techniques considering user preferences and traits (such as demography), and item similarities [6]. The main approaches for building recommenders are Collaborative Filtering, Content-based and Hybrid Systems.

For defining a recommendation task, let us assume that $D(u)$ is the set of items rated by user u, and $\overline{D(u)}$ is the complement of $D(u)$ denoting the set of items not yet seen by user u. Similarly, $U(d)$ is the set of users who have rated item d, and its complement is $\overline{U(d)}$. As defined in [6], the purpose of recommendation systems is estimating user preferences towards previously unseen items, which can be expressed by following:

$$\forall u \in U,\ \forall d \in \overline{D(u)},\ d' = argmax\ \omega(u,d), \tag{2.1}$$

where U is the set of all users, D is set of size $|D|$ of all possible items/documents, ω is a function representing "usefulness" of the item d for the user u.

In practical applications, ω can be represented as a utility matrix storing users' ratings of items D. For instance, user preferences can be defined on the scale from 1 to 5 with 5 being the highest rating as follows:

$$M = \left[\begin{array}{c|cccccc} & d_1 & d_2 & d_3 & d_4 & .. & d_n \\ \hline u_1 & 1 & 3 & 4 & & .. & 1 \\ u_2 & 5 & 3 & & 1 & .. & 5 \\ u_3 & & 3 & 5 & & .. & 1 \\ : & : & : & : & : & : & : \\ u_m & 2 & 3 & 5 & 1 & .. & \end{array}\right]$$

The recommendation goal is thus to predict blank cells (shown in blue) in the partially filled matrix $|U| * |D|$ of user ratings above.

Collaborative Filtering (CF). The CF recommendation technique is based on the assumption "if a person A has the same opinion as a person B on an issue, A is more likely to have B's opinion on a different issue than that of a randomly chosen person" [265]. The list of recommended items or products is based on the past ratings data from many users with similar tastes or purchase patterns [265, 139]. For instance, the Group Lens project employs collaborative filtering techniques for understanding user interests based on the direct input of users or based on the data stored in history logs [135].

CF is based on matching lists of preferences of users. When a user shares the same preferences or voting for same products, with a group of users, a new item, which was also preferred by the user group, can be recommended for the user. One of the drawbacks of the CF is the cold-start-problem, which requires knowledge on previous user activities to provide the user with recommendations. As a possible solution for dealing with rating sparsity and reducing rating prediction errors, Cai with co-authors [33] proposed a group-based approach rather than an item-based similarity of user ratings. They [33] create a "user-typicality" matrix for predicting user ratings based on user group similarities (users can belong to different groups) and compare their approach with another method of clustering, which associates users with their sole clusters [33], while referring to another work based on the matrix factorization technique and latent learning of user-item similarity in [284].

Nevertheless, an advantage of the CF is that it does not require knowledge about characteristics of the rated items since only user ratings are required [139]. Another important advantage of CF recommenders is the possibility to include user feedback (purchases), taking the temporal dynamics of changes in user taste into account. The CF works with a set of items represented by their identifying numbers, and the users' votes about these items. The CF approach can be formulated as follows [6]:

$$r_{u,d} = aggr_{\forall u' \in U(d)}[r_(u', d)], \tag{2.2}$$

where u' denotes users similar to user u and rated the item d, and $aggr$ is the aggregation function, which could be is calculated (in its simplest form) as an average rating of item d based on ratings of similar users u'.

Content-based Systems. In contrast to CF, the content-based approach builds user profiles based on a list of items preferred by the user in the past [6], without considering other users' preferences. One of the drawbacks of the Content-based RS is the situation when there are limited ratings or no previous ratings for the new user exist [139]. Another issue is the limited novelty (serendipity) problem [139]. The content-based approach can be formulated as [6]:

$$r_{u,d} = aggr_{\forall d' \in D(u)}[r_(u, d')], \tag{2.3}$$

where d' denotes items similar to the item d and thus the aggregation function $aggr$ can be calculated as an average rating of similar items rated by the user u.

Hybrid and Social-based Systems. Previous research work on personalisation and adaptive systems has exploited information published in social network platforms to collect information on user traits and interests. For instance, Abel with co-authors [2] uses Twitter to create content-based user profiles, which are further aligned with news articles in their news recommendation experiments. Abel with co-authors [3] demonstrated that information from several social networks, including Twitter, Facebook and LinkedIn can be used to provide improved recommendations.

As stated in [238], exploitation of different parameters and data sources should be considered in building state-of-the-art recommendation systems. Combining several techniques usually results in better recommendation performance [139]. The hybrid systems usually constitute a combination of content-based and collaborative filtering approaches [6]. In [102], an adaptive hypermedia system framework based on Twitter data was created to investigate different recommendation techniques by bringing several information sources from Twitter. The proposed hybrid recommendation technique uses not only user context, but also context shared in the user's social network. In a sense, this approach merges collaborative filtering and content-based approaches. In [101] authors suggest employing tweets' content extracted from the network of the user followers and friends. Using social network data was also discussed in [139].

It seems, however, that the generic adaptive framework does not address the possibility of user involvement to control the adaptation or recommendation process as advised in [258]. For instance, an option of selecting appropriate content might be useful when a user is not interested in the content generated by one's followers.

Other Techniques. Table 2.1 shows examples of the aforementioned recommendation system types, their benefits, and drawbacks. Other recommendation techniques include demographic, Utility-Based (UB) and Knowledge-Based (KB) [32]. Demographic Recommendation (DR) systems use data on user demographic traits, such as age and gender, to categorise users and establish rules for creating recommendations in accord with the group preferences [32]. The main benefit of DR is that it does not require previous information on user preferences since DR works with "stereotypes". The UB recommenders rank items according to their "usefulness", which is computed for each particular user based on item properties, which should satisfy user requirements. The KB recommends items based on knowledge of how specific things could satisfy specific users' requirements [32].

Algorithm/Technique	Example	Benefits	Drawbacks
Collaborative Filtering			
Latent Factor Models (Matrix Factorization Models), Time-aware Models, Nearest Neighbour, Clustering, Graph theory, Linear Regression, probabilistic Models, Bayesian Networks, Artificial Neural Networks	Recommending music at Last.fm, Group Lens project [135].	Possibility to include user feedback (purchases), consider temporal dynamics of user taste changes	Cold-start-problem when no previous user activities are known to the system (new user) [139].
Content-based			
TF-IDF Models, Clustering, Bayesian Classifiers, Decision Trees, Artificial Neural Networks	Book recommendations using text categorization [170]	No need of information on other users, Possibility to recommend new items not yet rated by others, Explicit Knowledge of the Content Features	Content Analysis Limitations, Cold-start problem for new users having little or no previous ratings, Limited novelty (serendipity problem)
Context-aware Systems			
Patterns discovery, data mining, information retrieval	Personalising to user location such as in COMPASS, tourist supporting IS, which provides recommendations based on user interests in sightseeing or other activity and based on tourist location [253].	Adapting recommendations to the specific user traits and needs.	Caution is required while selecting context-variables, which might negatively effect recommendation outcome [253].
Hybrid Systems			
Combining different techniques	The Entree system [32]	Can use the benefits of the combined approaches, allows a better recommendation performance [32]	The hybrid systems are designed to overcome the disadvantages of the combined methods.

Table 2.1: Recommender systems and techniques

Assessing Recommender Performance. The main evaluation metrics used for assessing recommenders' performance can be classified into accuracy-based metrics, non-accuracy metrics and user satisfaction metrics [107]. The information retrieval metrics such as accuracy, precision, and recall are widely adopted for evaluating recommendation performance based on comparing the predicted and actual ratings assigned by users. For instance, Mooney and Roy [170] exploit the information retrieval metrics for assessing content-based recommendations. The accuracy-based performance assessment can be executed quite efficiently on several data sets and algorithms in offline tests [107]. The selection of accuracy metrics in offline experiments is discussed in [94].

However, offline tests cannot measure user perception and satisfaction with the recommendations provided. This is why online user experiments can be performed based on explicit user feedback or with the help of implicit observations of user activities (such as purchases) with the system presented in [107]. Non-accuracy metrics such as the novelty of the recommended items and the amount or coverage of items recommended can be used when users are to be surprised with new items, which might be interesting to consider rather than focusing on accuracy of predicting very similar items to those seen before [107].

Furthermore, recommendation approaches should allow a balance between the quality of the provided recommendations and their novelty for the users [107]. Domain features and circumstances in which the user works with the recommender are important factors to keep in mind. Other factors playing an essential role in the recommendation process include the availability of relevant data on user traits and item characteristics. The possibility of dynamic changes in user references cannot be disregarded.

Considering CF algorithms evaluation, Herlocker with co-authors [107] discuss challenges of evaluating recommendation outcomes. They argue that the evaluation methods should be appropriate for the main purpose and nature of the recommendations. Overall, they suggest that the recommendation evaluation metrics and algorithms should be carefully selected while considering the expected recommendation outcomes, datasets' features, user expectations, rating scales, and sparsity. The size of the dataset, the number of users and items to recommend might impose constraints on the process of creating correlation tables. Some users might have very few available ratings in the data set. On the other hand, some recommended items might be more popular and have more ratings assigned compared to others. The dataset features and their distributions are thus paramount to consider when building offline evaluation experiments [107].

2.6 Research Gaps

To summarise, the above mentioned research points out cultural behaviour differences of users online and suggests culture-aware adaptive applications. It remains however unclear whether knowledge on user cultural background could help in improving recommendation performance as compared to country-related recommendations in particular application domains, such as in social networks and microblogs, in which user interests might extend beyond their country boundaries. Considering particular applications, networking friends such as studied in [38] and region-specific news or entertainment could be recommended.

2.6.1 Cold-start and Ratings Sparsity

Previous works including [33] and [284] address ratings sparsity and cold-start problems in CF recommendation systems. They discuss latent learning of user preferences, which is typically solved with the help of matrix factorisation techniques. While Cai with co-authors [33] propose a group-based approach and calculate user ratings based on their associations with user groups having similar perceptions on the rated items, thus avoiding sparser matrices with user-item based ratings, they explain their method within the psychological context of similarities in user perceptions and how they "typically" comprehend certain concepts or items.

In [33] the authors contrast their approach with the clustering-based methods, which are based on grouping or stereotyping of user preferences. However, they do not consider the homophily of user behaviour concerning their cultural preferences.

As was discussed in [162], homophily in different aspects of human life, which can be social status, gender-related roles or education, influence the formation of social relationships and human behaviour within different human groups and nations. When considering Twitter research, [11] observed a significant influence of age and political views when including the information extracted out of the social connections of the Twitter users.

2.6.2 User Traits and Context Mining

Since the primary goal of this thesis is to consider user cultural context in adaptive applications, we analysed research papers and recommendation techniques taking into account user contextual information. The main purpose was to find out the prominent techniques, their benefits, opportunities, challenges and pitfalls which needed to be further addressed. Since mining user-related data can threaten user privacy and lead to severe consequences in real life, we also need to refer to related research such a user profiling attacks and means of preserving user privacy online [5].

2.6.3 Social Links Formation

There are few previous works investigating the cultural differences of user online behaviour [142, 237, 80, 4]. Friendship network formation is studied in [142, 237] and demonstrated geographic locations and languages influence on the creation of the network connections online. Interestingly, flight connections and Twitter friendship connections are significantly correlated as discussed in [237]. However, there were no clear indications of the other aspects playing a role in the interpersonal communications online.

2.6.4 Time-frame and Features Analysed

In [80], Pearson correlation between cultural dimensions and aggregated online behaviour from the thirty top Twitter countries was analysed. The sociological models are exploited in [80] finding a strong correlation between some of the cultural dimen-

[5] Chapter 3 discusses mining user content and privacy threats in Twitter

sions and user behaviour on Twitter, particularly, user mentions, status updates and posting time. However, the set of the Twitter features is quite limited and requires further analysis given its relation to cultural behaviour patterns. It is also possible that the selected features were reflecting online user activities also correlate with events occurring in user locations during the period of data collection. This is why it is important to increase the features set and also perform analysis for a longer time frame to minimise the influence of real-life activities. Also, the authors mention the possibility of high variability in some individual microblogging behaviour. However, it is not clear how the behavioural outliers are dealt with and what authors suggest for treating such individual differences while building practical web applications.

2.6.5 Sample Size Limitations

Content analysis and statistical tests are used in [4]. They study microblogging differences between Japanese and American students, having more personal messages and questions respectively. The findings are discussed with respect to the sociological studies of Hall and Hofstede. The authors used content analysis of a relatively small sample (20 manually labelled tweets per user for 200 users) derived with the help of Twitter Search. They mention possible biases in the tweets analysis. An application of the proposed technique (using content analysis) to the larger dataset is not discussed and might be unfeasible in practice.

2.6.6 User Behaviour and System Design Boundaries

In e-learning applications such as facilitating learning collaborations, [283] suggest considering users' properties in the aspects of knowledge needs and competencies, social and technical preferences. The knowledge aspect reflects user interests and the level of competencies in some subject areas (for instance, mathematical skills prerequisites for attending some courses), which are matched for finding e-learning collaborators. The social aspect is discussed in relation to building effective communication networks and influencing other users. While the technical aspect is related to the users' technical preferences such as the use of media and communication synchronicity needs for selecting web services and applications [283]

We might argue, however, that user interaction styles and technical preferences in using applications might reciprocally influence each other. Firstly, the cultural user backgrounds can be reflected in user social interactions on the web as discussed in [81, 121]. Secondly, social networks and web services' features might impose limitations or restrictions on how users access and exploit them. This is why, while studying microblogs, we focus mainly on Twitter and avoid its direct comparison with other microblogging systems.

2.6.7 Exploiting Cultural Differences

Findings on cultural differences in user behaviour online can be further exploited for building state-of-the-art online communication tools and friend recommenders. However, the effectiveness of the culture-aware recommendations is yet to be investigated.

2.7 Conclusion and Discussion

In this chapter, we briefly outlined the related research in the fields of social informatics and microblogs exploitation for mining user context-related information required for the development of state-of-the-art applications such as recommenders. Despite the microblogging limitations, such as difficulties in mining data out of short and informal text messages and not accurate metadata, user-generated content can be used to extract user interests, geographic locations, and languages amongst other contextual information. Therefore, social networks can be used to address the "Cold-start problem" when user information is not explicitly available.

As explained, in the light of existing sociological research, distinctive behavioural patterns of users online behaviour, including user-follower relationships and usage of Twitter-specific features, can be used as a proxy for a country or cultural group predictions. Cultural stereotypes are however referred to specific countries or nations in the sociological studies exploited. Such simplifications might account for the very restricted use of the sociological models in practice such as critically discussed in

[163]. The assumptions on the heterogeneity of personal behaviour within country borders might be too simple to account for political or societal changes occurring in particular countries and globally. Also, it is essential to be aware of research biases and other relevant variables, which might be omitted from these analyses [111].

Personality traits across cultures in respect to particular geographic locations were studied in [15]. Moreover, Lewis [148] stated that Spanish people coming from particular regions might behave very similarly to linear-active people in the sense of productivity. The diversity of online interaction of Spanish and Dutch social networking users was studied in [12] finding out different communication preferences in respect to Hofstede's cultural dimensions.

Previous research works suggest the possibility of exploiting user country information in adaptive applications such as recommendation engines. Practical realisations might include urban planning, communication tools responsive to cultural user preferences and friend recommendations. However, we still require a thoughtful assessment of the performance of such culture-aware adaptations. The online communication of users in social networks requires a further investigation into how the different cultural groups interact, what are the most important features to consider for predicting user activities online, which features could provide cues on user origins and to what extent user origins could be mined out of user-generated content and activities.

In this review, we focused on the social web, user-context extraction out of social networks and microblogs for further exploitation in state-of-the-art web applications. The ethical issues related to user privacy we address in the next chapter, in which we also outline technical details of data collection and mining user traits out of microblogging content and metadata.

Chapter 3

Mining User Content and Privacy

"Very private people have mastered the art of telling you little about themselves but doing it in such a way you think you know a lot."

- Anonymous at https://www.goodreads.com/quotes/tag/privacy

3.1 Introduction

Internet technologies make it possible to store and transfer large volumes of data. These data often include cues on user online behaviour and personal information, collected and processed by web services and applications. For instance, social websites such as Twitter and related web/mobile applications allow their users to connect with friends, sharing information in real time. Moreover, microblogs and other online resources can also be exploited in business settings for marketing and research purposes. However, the microbloggers' personal privacy needs to be observed and weighted against the practical benefits provided by microblogs.

Online users' privacy could be supported with the help of a regulatory framework and software controls. In support of the human right for privacy, there are national and international regulations being developed to preserve personal privacy in an online setting. In this respect, the Organisation for Economic Co-operation and Development guidelines [178] address online privacy protection and safe information transfer via computer networks in order to prevent unlawful personal data access, storage, and processing. In support of the human right for privacy and their users' benefit, SN websites and applications provide functionality to exercise privacy control of the shared user-generated content and metadata. There are different user profile settings and options across SN websites helping users to hide sensitive information.

Software means and privacy settings are not entirely effective yet. Privacy policies are not practically useful since personal information can be derived out of social networks with information retrieval, data mining and machine learning methods [52] or named-entity recognition techniques [277, 157]. Also, the privacy of sensitive micro-posts can be violated by the user's friends reposting initially protected content [164]. Mentioning users and adding their geographic locations is a possible threat of violating user privacy despite imposed rules and overall satisfactory design of Twitter privacy controls [188]. Some Twitter software clients facilitate information leakage from protected users [164]. Sensitive topics on diseases or alcohol consumption could also be revealed by Twitter users, which could benefit from software assisting tools in protecting users from posting sensitive messages [157].

In this chapter we outline the main issues and means of protecting user privacy online with a focus on SN and Twitter microblogs. Besides, we look into the ethical considerations of researching openly-available user-generated content published in Twitter.

3.2 Privacy Protection in Twitter

Twitter would be much less useful for sharing news and information in real-time and finding users with similar interests online if everyone on Twitter protected their status updates. Protecting Twitter messages might be seen as counterproductive for reaping business opportunities online. However, in some cases, users opt to protect their tweets to safeguard their personal data. In this section, we discuss the privacy protection on Twitter and other social networking applications, possible threats, and solutions. To provide the context and introduce Twitter terms, we start by describing the main features and privacy controls in Twitter.

3.2.1 Twitter Microblogs

Short Messages, Links and Hashtags. Initially, Twitter microblogging platform allowed its users to post short messages of not more than 140 characters for sharing live events, news and interesting and useful links. In 2017, Twitter has doubled its character limit to provide more space for users microblogging in English, and Twitter experiment on a small number of users showed that the number of abandoned (when hitting the text string size limitation) tweets decreased from 9% to 1% [204], while microbloggers tweeting in Japanese, Korean, and Chinese would continue with 140 characters limit due to the higher expressiveness of the glyphs.

To accommodate for the text length limitations in the previous Twitter versions and for further statistics over URL access (the shortening services might provide additional information on a referrer, its software and hardware parameters), users might exploit URL shortening services such as bit.ly. Hashtags are Twitter "keywords" and begin with a "♯" symbol and allow tweets to be labelled and organised into trending topics, which could be used by event organisers to group event-related tweets and analyse user opinions and feedback. For instance, conference attendees might use pre-defined hashtags for tagging their messages for organising all the related messages and resources.

Retweets, Mentions and Replies. When a particular tweet is of interest for the user, the tweet can be "retweeted", which enables tweets to spread over the follower network of the person who "retweeted" the tweet. The retweet is similar to the "forwarding" feature in e-mail systems. Starting from 2016, it is possible to retweet one's tweets to gain more attention to a previously published message. As described in [116], Retweet messages usually begin with the "RT" symbols can be considered as an "old school" manner of sharing retweets, a "convention" initially used by users. Retweeted Twitter messages often also included "via @author" at the end, in which the author denoted the initial author of the tweet. The "old school" retweet had a benefit of adding own content to the original message, and in 2009 Twitter introduced own's "Native Retweet" with an added button into the user interface; with this feature users could not edit the initial retweet, however, authors of the original messages were automatically acknowledged [116]. It is important to mention that currently, it is possible to add an own comment to the retweeted message. Usually, a "@" symbol denotes the username of a Twitter user mentioned in the tweet. Twitter replies start from the "@" and enable users to form a dialog, such as in chat rooms. In 2017 Twitter excluded user mentions from the tweet length limitation (mentioned users) and help to de-clutter messages from the list of mentioned users (they are shown above the tweet) for improved user experience [174]. Users who follow each other can send direct messages, which can be compared to e-mail messages and cannot be searched in Twitter.

Friends and Followers. Influential users. The messages, or status updates, become visible to users who follow their authors. Users see their messages and the messages of their friends. The messages of the followers are typically not seen on the user home page unless the user is mentioned using the "@" symbol. This way users can choose which messages they see on their home page. For simplicity reasons, we assume that the followers to friends (following) ratio can be used to assess the "influence" of a particular user. However, as discussed in [35], in-degree influence is only one measure of influence in Twitter. The most mentioned users are typically celebrities and public persons and the most retweeted content providers influence microblog readers [35]. Additionally, [35] suggest that the number of tweets and friends (followers) can be used to identify electronic agents and spammer scripts.

Privacy Controls in Twitter. Overall, despite some user-related information leakage, Twitter can be seen as providing quite reliable privacy mechanisms [188]. Twitter privacy controls, however, can ensure only privacy of sharing information at the specific time when these privacy settings are enabled. Users might disable their tweet locations, tweets, and accounts from access by the general public. In the case of protected, or closed user profiles, only user followers can see the user's tweets and profiles. However, to follow the protected account, a request should be approved by the owner of the protected account. Followers can thus access their friends' accounts, tweets, retweets, and favourites. In [188], Twitter[1] privacy design choices, conditions and rules of information sharing are formally described.

3.2.2 What is Personal and Sensitive Data?

For protecting people's privacy rights, personal data transferred over the Network should not be compromised. Therefore, it is important to identify which personal data might be considered as sensitive, requiring secure handling. Following the privacy guidelines outlined in [159], personal data relates to information which could be used to directly identify persons and includes, but is not limited to, the following:

- First and last name of the person or their immediate relatives;
- Birthplace and date;
- Physical address;
- Financial records;
- Medical records;
- Employment details.

Moreover, it is necessary not to underestimate the importance of the user's contextual information, which could provide additional cues for identifying personalities and which potentially could affect their privacy. For instance, some of the information such as a user's Internet Protocol (IP) address might be considered as not personally-identifying information; however, the IP address could still be used to identify a specific user when having access to the information of the related user ID (or login information) [159]. Mining information from microblogs might be used to identify particular users, which raises privacy concerns in some instances, such as when sharing personal health information [224].

[1] Facebook's privacy is also discussed

3.2.3 Transparency and User Control

Social networking websites such as Facebook, Twitter, and LinkedIn often maintain user data including a user's network connections, geographic locations and even employment positions respectively, when provided by the user. Some web applications collect user personal details explicitly or collect their online activities with user permission [212]. However, some of the websites do not inform their users about data collection taking place at the time of their visit [212]. To ensure that user personal details are not shared unwillingly, users should exercise complete control over their personal details. The consequences of sharing the personal data should be thoroughly thought through by the users, who ideally should have complete control over their data [108]. As a possible solution, Papadopoulou with co-authors [187] suggest to exploit personal data stores, in which users manage their privacy settings while remaining sole owners of their data. Service providers could further adjust their services in respect to available user data and associated privacy settings [187].

3.2.4 Data Mining and Privacy in Twitter

Privacy is one of the most important factors influencing microblogs' adoption by users according to a study by [88]. Microblogs might be perceived as unsafe compared to other social platforms and therefore not exploited by some users who are cautious about sharing their personal information online [88]. Indeed, the social web makes it possible to mine openly available user data [108]. Users might publicly share their preferences or user activities, and traits could be automatically mined based on their online behaviour [108]. As a result, user behaviour patterns could be exploited to deduce user specific traits. For instance, user geographic location could be inferred based on user-generated content, metadata associated with microblogs and user social networks (as we demonstrate in the following part and in [121, 52]). Personal information can also be derived out of user social networks and online web resources with the help of data mining and named entity recognition techniques [277]. Thus, explicit user personal information sharing is not required to collect potentially sensitive information [108].

Table 3.1 (on the next page) summarises the application opportunities and challenges of data mining in Twitter microblogs. The classification of data types is adopted from the work by Liao with co-authors [151]. The only difference is that we distinguish between micro post's contents and associated metadata. It is important to mention that several data types and sources can be combined and different techniques applied to achieve or evaluate the inference results. For instance, Alex with co-authors [13] use gazetteers to infer locations defined in Twitter profile, which accuracy is assessed with the help of human assessors and Google Maps applied to tweet's content.

While analyzing Twitter content, an application of clustering methods including K-means and hierarchical clustering was studied in [95]. Authors collected tweets voicing opinions and attitudes of Twitter users towards specific countries and sport players in World Cup 2015 events. The study demonstrated the better performance of K-means algorithm for the more massive datasets as compared with the hierarchical clustering approach [95].

A hierarchical cluster analysis was used to investigate human mobility patterns for top twenty tourist locations reflected in Twitter data with geolocations attached [19]. Authors emphasise the usage of big data useful for users mobility studies, which could benefit from the clustering methods based on the openly available content. They ranked touristic sites locations in respect to user residence locations based on heuristic rule considering at least third of the most frequent tweeted locations. Interestingly, Eiffel tower was amongst the five top visited locations [19].

A temporal coherence of tweeting behaviour was exploited for analysis of urban land usage in City of Chicago in [229]. Authors investigated the time of tweeting at top user locations, which are associated with spatial clusters of tweets. The usage of DBSCAN algorithm enabled to abstract from the defined prior knowledge on the number of clusters [229].

Data	Algorithm/Technique	Application Opportunities	Drawbacks and Challenges
User Properties			
Location, user name, user description and language fields	Content analysis, named entity recognition [13], text classification with Multinomial Naïve Bayes [104],	User location detection [52, 104, 13]	Reliability of user information, which can be obfuscated to protect user privacy [151]
Social Connections: Social Links and Interactions			
Friendship connections between users, and their interactions such as Retweets, Replies and Direct Messages	Pattern mining, interaction analysis of social networking sites (authors do not specify however if Twitter media was used, they also exploit geographic locations extracted from metadata) using Latent Class Models [12]	Finding communities and relationships, content propagation analysis; user preferences towards social networking usage in respect to cultural traits [12]	Connections change over time
Message Text			
Free-text of the Tweet's message	Bag of words technique to represent unstructured text, TF-IDF Models to consider importance of the words, LLM; K-means and hierarchical clustering [95]	User opinion mining such as towards sport events and countries voiced [95], location detection (using LLM)	natural language processing needs [95], challenges of mining short text messages, often including abbreviations and wrong spelling [120]
Message Metadata			
Tweeting Global Positioning System (GPS) location coordinates and geographic place when available, time of posting, hashtags, web page addresses and user mentions included into the text message	Geo-location, Pattern discovery, usage time pattern mining; hierarchical cluster analysis [19]	Trend detection using hashtags, finding user locations, analysing behavioural patterns [121]; sightseeing mobility patterns across country clusters [19]	Location data is often missing or inaccurate, need for human involvement to mark specific points of interests [19]

Table 3.1: Data mining techniques applied to Twitter data

3.2.5 Information Leaks

The Twitter microblogging platform enables users to protect their accounts. However, due to the nature of microblogs and their attractive openness, only a small fraction of about 10% of users protected their accounts in [200]. It is not trivial to manage private information amongst friendship networks [92]. Indeed, even protected tweets can become publicly visible when being reposted by the user's friends [164]. In [164] dataset, about 1% of all accounts retweeted private information of their contacts. A study into how personal information can be revealed in microblogs was reported in [34], exploiting machine learning and human annotations to estimate the level of sharing of personal information by microbloggers. They suggest a "privacy score" which could be applied to user contacts to help in decision making on how much information could be shared with these contacts [34]. As Grimmelmann [92] pointed out, existing privacy issues do not stop SN users willing to communicate online with their friends and acquaintances.

As one of the solutions for supporting user privacy, McCallister et al. [159] suggest using anonymizer methods to avoid personal information leakage. A prototype using cryptography and access control of followers was proposed by De Cristofaro with co-authors [57] to protect user-generated content including hashtags. In support of personal data protection, various methods of information coding were developed such as proposed by Bayardo and Agrawal [20]. However, the coded information usage may be detrimental to information usage or affect the quality of the stored information [20]. It might be inconvenient for using encrypted information in practical applications since social networking websites are specially built to enable user communication and sharing content. Thus, usage of these applications and services requires users to share information to a certain extent, allowing them to exploit website functionality as they need.

3.3 Research Gaps

3.3.1 Privacy Protection in Microblogs

To summarise, Twitter microblogs allow users to share their status updates and interesting links. For personal or corporate use, the majority of Twitter accounts are open. Protected accounts are prone to information leakage and require further care when sensitive or personal information is shared within friendship networks. It seems, however, that protecting tweets might be counter-productive and unnecessary since protected profiles might block users from new business opportunities in detriment of fostering user communication on Twitter. In chapter 8, we further analyse whether protected users communicate and form relations less intensively compared with open users and might, therefore, take less advantage of the Twitter messaging service.

3.3.2 Privacy Need and User Origins

When referring to the human rights regulations, Whitman [261] discusses perceptions of privacy in different cultural settings. It seems that societal values impact the view of privacy for different nations [261]. In [157] authors examine sensitive information leaks and found differences in the information leaks, particularly on the topic of "depression" for Singapore, the UK, and the US. The level and type of personal information sharing differ for users from the USA and Singapore, in which persons from States tend to reveal more personal information (contact, demographic, education, and job), while Singaporean persons mostly share their feelings and attitudes [59]. It is reasonable to assume that the privacy settings could be used differently amongst cultures. This is why it is interesting to investigate further how privacy controls are exploited by different cultural groups in Twitter.

3.4 Conclusion

In this chapter, we described the Twitter microblogging platform and related research. Microblogging content and metadata provide vast opportunities for extracting user preferences, traits, and interests, however, user privacy needs should not be neglected, especially when we consider different user needs in respect to user cultures. Would protecting user accounts be counter-productive and unnecessary? In chapter 8, we will address this controversy. Next, we describe term of culture and outline sociological research related to this thesis.

Chapter 4

Models of Culture

> "Culture is like gravity: you do not experience it until you jump six feet into the air"
>
> - F.Trompenaars [241]

4.1 Introduction

In this section, we define the term of "Culture" for usage in the context of this research. We analyse the application of sociological models in Web information systems and networking research fields, describe their commonalities and differences and conclude on selecting a model to study microblogging behaviour in Twitter.

4.2 Culture and Nationality

If we look into an English dictionary such as provided by Oxford[1], we find out quite a few definitions of word "Culture". Culture is used in the context of arts and human intellectual skills, can be viewed from an anthropological point of view when studying cultural groups' attitudes and customs, and is applied in biology and agriculture when referred to the growth of bacteria or plants.

In the study of user behaviour online, we exploit the term of "Culture" in an anthropological context. We are interested in understanding how users from particular

[1] http://www.oxforddictionaries.com/definition/english/culture

countries and regions exploit microblogging features, and what consequences these differences have which could be practically exploited in social networking settings and web applications. As we will see next, sociological models such as Hofstede and Lewis exploit the term of culture when referring to persons coming from particular nationalities. In [110] Hofstede defines Culture as:

Definition 1 *Culture is the collective programming of the mind which distinguishes the members of one category of people from another.*

In his study of religion Clifford Geertz [84] defines culture as:

Definition 2 *A historically transmitted pattern of meanings embodied in symbols, a system of inherited conceptions expressed in symbolic forms by means of which men communicate, perpetuate, and develop their knowledge about and attitudes toward life.*

Kroeber's definition provided by The Center for Advanced Research on Language Acquisition[2]:

Definition 3 *Culture consists of patterns, explicit and implicit, of and for behaviour acquired and transmitted by symbols, constituting the distinctive achievements of human groups, including their embodiment in artefacts; the essential core of culture consists of traditional (i.e. historically derived and selected) ideas and especially their attached values; culture systems may, on the one hand, be considered as products of action, and on the other as conditioning elements of further action.*

For the purpose of studying user behaviour online, we could define a simplified term of Culture by nationality. It is important to mention, that we group users from particular countries and refer to these groups as "Cultural groups", "Cultural dimensions", "Cultures" or origins, which we exploit interchangeably with the following meaning:

Definition 4 *A cultural group is a group of people who share a common language, history, social habits, customs and geographic regions.*

[2] `http://www.carla.umn.edu/culture/definitions.html`

When considering an application to the social networking behaviour of particular user groups, we could further adjust this definition as follows:

Definition 5 *A cultural group is a group of people sharing a subset of languages, geographic regions and infrastructures at place, societal values, lifestyle preferences and communication patterns.*

The "subset of languages" considers multi-lingual cultural regions in which different sub-cultures share a set of different languages or dialects. As an example, Welsh is widely spoken in the north and west of Wales [3]. Therefore, people in this regions might prefer the Welsh language over English in their daily life. However, for our thesis and to adhere to the previous sociological model by Lewis [148], Welsh and English speaking persons from other parts of England are included in the same culture group of "British", in the "Linear-active" dimension [4]. Lewis emphasises that "The majority of British people bear little resemblance to the stereotype. Not only is the image one of an upper-class personage of a former era, but it does not take into account regional differences, which in the U.K. are extremely marked" [148]. It is important to mention that we do not aim to stereotype persons in accord with the definition provided and the previous sociological research. Instead, we use these concepts as a starting point for referring to the defined human groups sharing a similar social living environment. In this sense, we simplify that generally British people have more similar attitudes and communication styles as compared to other cultural groups as German. Also, behavioural differences should be even more pronounced in South European cultures which generally are perceived as more affectionate [148].

Throughout this thesis and while referring to microblogging behaviour, we follow this last definition, assuming that user communication in microblogs is influenced not only by real-life events, but also their communication preferences originating from their cultural backgrounds and social surroundings. Such an understanding of

[3] https://en.wikipedia.org/wiki/Welsh_language

[4] Our human coders agreed that English-speaking and Welsh-speaking persons belong to the "Linear-active" users group in accord with the Lewis Model of Cultures, please refer to Table E.1 on page 260 in Appendices.

general preferences for particular cultural/geographic origins could provide valuable information for adapting web applications and software to specific user needs, e.g. when such information is required, and could be used to improve user experience further online.

4.3 Applications of Models of Cultures

We critically review existing sociological models and their applications in information systems (IS) research and industry. We coded cultural dimensions (or factors) for each model in brackets, prefixed with the initials of the models' first authors. For instance, the cultural dimensions above start with the "GH_" denoting that Geert-Jan Hofstede developed the model (referring to his works [109, 112]).

4.3.1 Sociological Models Compared

Herein we refer to the sociological models of culture referred in IS research, which include works of G.J. Hofstede [109] and Fons Trompenaars [241], but also the lesser-known Model of Cultures by Lewis [148]. A systematic way is needed to understand better cultural differences of persons from different nations to improve managerial practices in multinational organisations operating in global economies, influenced by human socio-cultural factors [109]. In [109, 112], IBM employees from more than 50 countries were interviewed and scored following the defined cultural dimensions reflecting cultural behavioural preferences in accordance with Hofstedes Model including:

- Power Distance (GH_PD) defines the extent to which power (power, or authority, here reflects perceptions of authority figures or persons standing on the "higher hierarchy" level as they are seen by "lower" rank individuals, for instance, within nations, cultural groups or organisations wherein government or managerial figures can be seen as more "authoritative" as compared to their subordinates) is perceived in the society, for instance, a lower GH_PD indicates that power is distributed equally, while higher GH_PD indicates that society is very hierarchical in its functioning;

- Individualism (GH_ID) defines the extent to which the personal is given more importance as compared to collective, for instance, low individualism score indicates that the cultural group is more predisposed towards forming collective behaviour, while high individualism cultures are focused on individual personal needs;
- Masculinity (GH_ML) defines the cultural values of achievement and material rewards as opposed to "Femininity", which values more cooperation and better quality of life;
- Long-term Orientation (GH_LT) defines the cultural views on traditional values versus adaptation in favour of future well-being, therefore higher scoring GH_LT tend to adapt to the newly occurred circumstances to tackle fast-paced challenges in time;
- Uncertainty Avoidance (GH_UA) defines the cultural predisposition to disregard different thoughts which might be perceived as uncertain and therefore risky, low scored GH_UA cultures tend to accept new ideas easier than highly scored GH_UA cultures;
- Life Indulgence (GH_LI) related to cultural preferences towards enjoying various life aspects, for instance, high GH_LI countries prefer to be in control of their happiness when enjoying the life, while low scoring LI countries tend to adhere to stricter social norms.

Trompenaars' work [241] is based on a questionnaire filled out by 30 thousand respondents from 30 countries. Participants were mostly managers, assessed in respect of several cultural dimensions such as "Universalism vs. Particularism", "Individualism vs. Communitarims", "Neutral vs. Emotional", "Specific vs. Diffuse" ("specific" means here prescribed by protocol), "Achievement vs. ascription", and attitudes towards time and environment. Findings reveal that business organisation, its purpose and establishment of rules depend largely on the perceptions of employees influenced by their cultural backgrounds. Moreover, some countries are inherently multinational and diverse. This is why strong nation-wide stereotyping might not be appropriate. Personal attitude also differs among nations, age and gender groups. The cultural dimensions developed by Fon Trompenaars together with Charles Hampden-Turner are listed below [241, 165, 166]:

- Universalism versus Particularism (FT_UN) denotes an extent to which the rules and procedures versus personal circumstances take precedence in the decision making process. Cultures scored “high” in the FT_UN dimension prefer to adhere to more objective procedures and rules, while cultures scored “low” in the Universalism, or “high” in the Particularism, allow more flexibility by considering personal circumstances of people involved into the decision making process;

- Individualism versus Communitarianism (FT_ID)[5], in which cultures scored “high” in the Individualism FT_ID dimension value individual achievements and needs more over the group, cultures scored “low” in the FT_ID (and “high” in the Communitarianism) value more groups’ importance while avoiding mentioning individual achievements;

- Specific versus Diffuse (FT_SP) whereas specific cultures tend to separate personal and work lives, diffuse cultures tend to interweave business and personal life aspects;

- Neutral versus Emotional (FT_NE), cultures associated with the “Emotional” factor tend to show their emotions and strive to have a positive attitude, while “Neutral” cultures observe and effectively manage their emotions in general;

- Sequential versus Synchronous Time (FT_ST) in which sequential time management cultures aim to keep up to deadlines and generally do one thing at a time, while “synchronous” cultures allow more flexibility in time management, being able to work on several projects in parallel;

[5] we can observe a relevance with Hofstede’s GH_ID dimension

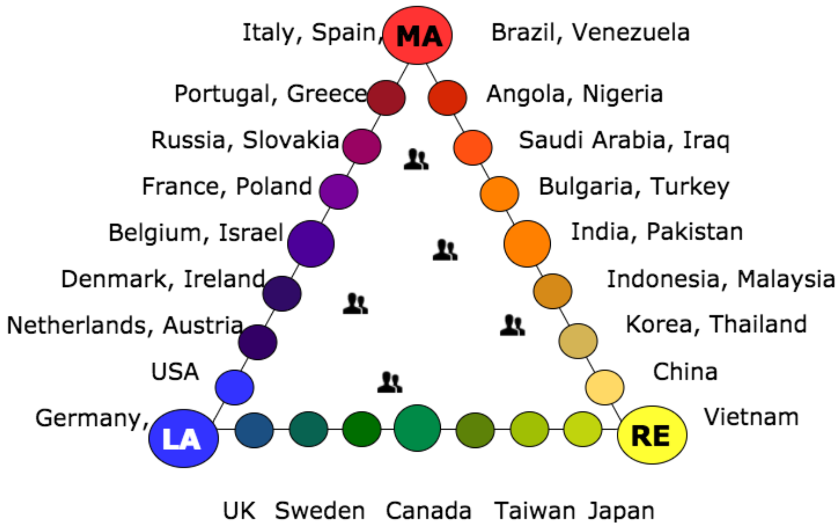
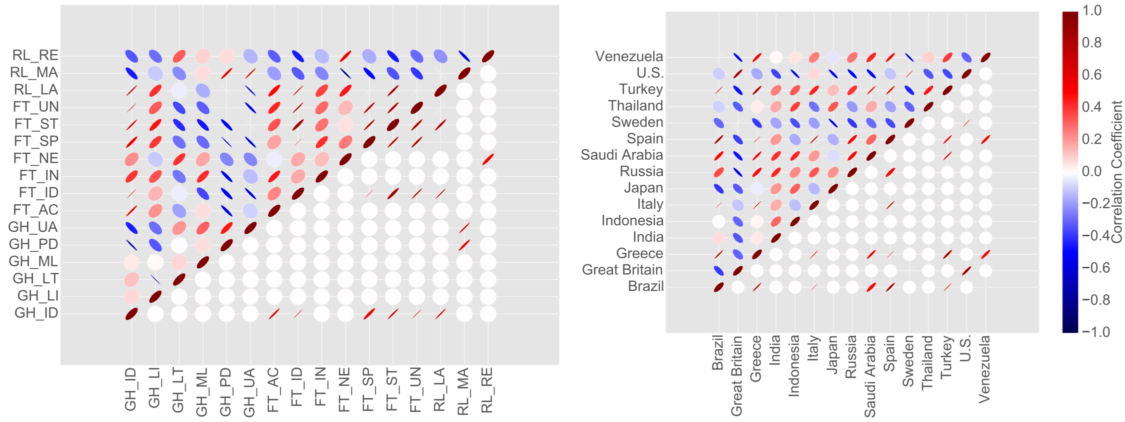


Figure 4.1: Simplified model of cultures by R. Lewis described in [148]

- Achievement versus Ascription (FT_AC) in which "Achievement" cultures value improved performance over position or authority such as in "Ascription" cultures;

- Internal versus Outer Direction (FT_IN) whereas "Internally-directed" cultures tend to take charge over their environments and allow constructive conflict, and "Outer-directed" cultures tend to avoid conflict situations while allowing to be controlled by their work environment.

The Lewis Model of Cultures is relatively less explored in the IS research, and relates cultural behaviour patterns with national groups. The model is represented in a shape of triangle showing the most extreme cultural profiles in the triangle apexes as shown in Figure 4.1. There are three cultural dimensions including Linear-Active (LA or RL_LA), characterising task-oriented and good planners such as persons from Germany, Multi-Active (MA or RL_MA), loquacious and people-oriented persons from Spain and Brazil, Reactive persons (RE or RL_RE) from Vietnam who tend to avoid conflicts and hide their emotions generally. The model links national groups with the defined behavioural profiles, however, it does not imply that individual characteristics must match [148] their national groups, instead, persons might have different mixtures of the cultural traits based on these three cultural dimensions.



(a) Cultural Dimensions (b) Countries

Figure 4.2: Relationships between cultural dimension scores and countries (after diagonals we see only significant correlations with $p - value < 0.05$ and Pearson correlation coefficient $r >= .52$ for n=15 observations; the initial (normalised) data is provided in the Appendices, Table B.2)

Based on insights and data published in [110, 241, 148, 158, 165] we summarised examples of selected country scores in the aforementioned sociological models and their assessment factors, called "cultural dimensions", in Table 4.2. Please note that we artificially categorised Hofstede cultural dimension scores into "High" and "Low" scored countries considering scores in 25th and 75th quantiles [6] respectively while regarding scores between as "Middle" scores. Cultural dimension scores of the Lewis Model of Cultures are derived from the approximate distances to the apexes of the model triangle[7]. We scaled scores of these three models to be in the same values range of [0..1] and calculated Pearson correlation for 15 selected countries, as shown in Figure 4.2.

Overall, it is interesting to note the overlap in the cultural models of Fon Trompenaars and G.J. Hofstede in respect of individual and collective attitudes in dimensions FT_ID and GH_ID respectively, amongst other strong correlations shown in Figure 4.2 (a). Several cultural dimensions correlate strongly across different models within selected country profiles. We also found that some countries have positive correlations in their cultural scores as seen in Figure 4.2 (b). For instance, Brazil

[6] Using Python pandas.qcut - quantile-based discretisation function with lambda x: pandas.qcut(x, 3, labels=['low','mid','high'])

[7] available at: http://www.crossculture.com/latest-news/the-lewis-model-dimensions-of-behaviour/

Model	ρ	significance level $p - value$
Hofstede	.11	.27
Trompenaars	.16	.11
Lewis	.31	.001

Table 4.1: Pearson correlation between flight distances (see Appendices for flight distances between countries, Table B.1) and distances between cultural dimension scores (see Table B.2 for used cultural dimension scores) for 15 aforementioned countries

cultural dimensions correlate significantly with cultural dimensions of Greece, Italy, Saudi Arabia, Spain, Turkey and Venezuela, while cultural profile scores of Great Britain correlate positively with Sweden and the United States. We must note the relatively small sample size, this is why we have shown only significant correlations with $r > 2/\sqrt(n)$ (n=15 is the number of countries considered) as a rule of thumb after diagonal of the correlation matrices in Figure 4.2.

Moreover, we found out that some of the cultural dimensions overlap in their respective geographic locations or countries. For instance, RL_LA overlaps with GH_PD_LOW, both including Great Britain, Sweden and the U.S., GH_PD_LOW overlaps with FT_IN_HIGH, both including Great Britain and U.S.

Interestingly, we found also positive relationships between bird flight distances and cultural dimension distances as seen in Table 4.1. The most significant relationship was found for the Lewis model dimensions. We could explain it by the inherited link of cultural psychological traits to national geographic locations [15].

Cultural Dimensions	"Low" Scored	"Middle" Scored	"High" Scored Countries
Hofstede's Model of Cultures [109]			
GH_ID	Greece, Indonesia, Saudi Arabia, Thailand, Venezuela	Brazil, India, Japan, Russia, Turkey	Great Britain, Italy, Spain, Sweden, U.S.
GH_LI	India, Indonesia, Italy, Japan, Russia	Greece, Saudi Arabia, Spain, Thailand, Turkey	Brazil, Great Britain, Sweden, U.S., Venezuela
GH_LT	Brazil, Saudi Arabia, Thailand, U.S., Venezuela	Great Britain, Greece, India, Spain, Turkey	Indonesia, Italy, Japan, Russia, Sweden
GH_ML	Russia, Spain, Sweden, Thailand, Turkey	Brazil, Greece, India, Indonesia, Saudi Arabia	Great Britain, Italy, Japan, U.S., Venezuela
GH_PD	Great Britain, Italy, Japan, Sweden, U.S.	Brazil, Greece, Spain, Thailand, Turkey	India, Indonesia, Russia, Saudi Arabia, Venezuela

Table 4.2: continued on the next page

Cultural Dimensions	"Low" Scored	"Middle" Scored	"High" Scored Countries
GH_UA	Great Britain, India, Indonesia, Sweden, U.S.	Brazil, Italy, Saudi Arabia, Thailand, Venezuela	Greece, Japan, Russia, Spain, Turkey
Trompenaars [241, 166, 165] [8]			
FT_AC	Indonesia, Saudi Arabia	Brazil, Great Britain, Greece, India, Italy, Japan, Russia, Spain, Sweden, Thailand, Turkey, Venezuela	U.S.
FT_ID	Japan, Thailand, Venezuela	Brazil, Greece, India, Indonesia, Italy, Russia, Saudi Arabia, Spain, Turkey	Great Britain, Sweden, U.S.
FT_IN	India, Russia, Saudi Arabia, Sweden	Brazil, Greece, Indonesia, Italy, Japan, Spain, Thailand, Turkey, Venezuela	Great Britain, U.S.
FT_NE	Brazil, Italy, Spain, Venezuela	Greece, India, Indonesia, Russia, Saudi Arabia, Sweden, Thailand, Turkey, U.S.	Great Britain, Japan
FT_SP	India, Japan, Russia, Spain, Venezuela	Brazil, Greece, Indonesia, Italy, Saudi Arabia, Thailand, Turkey	Great Britain, Sweden, U.S.
FT_ST	Japan	Brazil, Greece, India, Indonesia, Italy, Russia, Saudi Arabia, Spain, Thailand, Turkey, Venezuela	Great Britain, Sweden, U.S.
FT_UN	Indonesia, Japan, Russia, Venezuela	Brazil, Greece, India, Italy, Saudi Arabia, Spain, Thailand, Turkey	Great Britain, Sweden, U.S.
Lewis Model of Cultures [148] [9]			
RL_LA	Brazil, Greece, India, Indonesia, Italy, Japan, Russia, Saudi Arabia, Spain, Thailand, Turkey, Venezuela		Great Britain, Sweden, U.S.
RL_MA	Great Britain, Japan, Sweden, Thailand, U.S.	India, Indonesia, Turkey	Brazil, Greece, Italy, Russia, Saudi Arabia, Spain, Venezuela

Table 4.2: continued on the next page

[8] Since the model does not consider the "middle" level for specific dimensions, we assigned the "Middle score" to countries not included into the related dichotomous dimensions

[9] scores defined arbitrary when close to the Lewis' triangle apexes

Cultural Dimensions	"Low" Scored	"Middle" Scored	"High" Scored Countries
RL_RE	Brazil, Great Britain, Greece, Italy, Russia, Saudi Arabia, Spain, Sweden, U.S., Venezuela	India, Indonesia, Turkey	Japan, Thailand

Table 4.2: Cultural dimension scores for the selected countries

In short, we described cultural factors of the three prominent Models of Cultures. Some of these cultural factors are related to similar overlapping geographic locations, which correlate with each other significantly. Next, we are going to review how these cultural factors (or dimensions) are explored in web science or IS research.

4.3.2 Applications of Models of Cultures

Building upon the research in [109, 148, 158], we examined nationality importance in respect of political, social, psychological personal differences shaped during childhood age, and which could apply to web design considerations. To exploit these personal differences in e-commerce for building culture-aware websites, Marcus and Gould [158] emphasised the need for affordable web interfaces tailored to meet specific cultural preferences. Future work might, however, provide larger number of websites analysed, with a less subjective selection. Alexander with co-authors [14] used Hofstede's model to explain web interface preferences in web page layout, visual content and colour and proposed guidelines for creating cross-cultural usability design. Hofstede's cultural dimensions were also exploited for the analysis of music preferences across different countries for more than 53,000 Last.fm users [70]. They found a statistically significant correlation between cultural traits and music preferences, for instance, Long-term orientation (LT) and diversity in listened music genres, while countries with lower Life Indulgence indexes showed lower correlation with the music listening diversity metrics. In a situation of cold-start (when user is new in the system), they suggest to personalise music recommendations based on information about user countries, possibly exploiting social media websites including Twitter or Facebook [70]

Acar [4] compares usage of Twitter by American and Japanese students in respect of their cultural differences. They refer to cultural differences in respect to high-context and low-context differences such as described by the Hofstede and Hall's models [4]. Their findings reveal that American students post less personal messages compared to Japanese students, which refrain from referring to other persons. Acar [4] explains this by the "social sensitivity" of the Japanese individuals. Japanese users not only tend to avoid mentioning others but also ask fewer questions. In contrast to American students, Japanese students prefer to post more personal messages. The findings were achieved with help of content-analysis and statistical tests based on 4000 tweets published by 200 Twitter users (20 tweets of each user were analysed).

Twitter as a source of cultural behaviour traits explained using the Hofstede's cultural dimensions, and the study of Levine's Pace of Life theory was studied in [80]. Their experimental setup included the 30 top most active countries on Twitter. Their findings revealed strong negative correlation between cultural dimensions such as individualism and pace of life with user microblogging behaviour such as mentioning of other users and temporal predictability of user activities on Twitter respectively. Countries usually characterised as tolerant towards communication with power distant persons, also engage with microbloggers any in-degree friendship values. Also, aforementioned Twitter behavioural patterns correlated with the sociological and economic factors including Gross Domestic Product per capita, Education and Inequality [80]. Authors suggest to exploit found communication differences for people from different countries towards building effective communication tools and web applications such as friends recommendation services [80].

The work by Gao with co-authors [78] compares user behaviour on Twitter and Weibo, explaining the differences in temporal microblogging patterns, shared sentiment, hashtags and linking behaviour by the cultural differences in view of the Hofstede's research. The possible influence of the both system design characteristics was not considered in the analysis.

Hofstede's model is widely applied in studies comparing social networking with the help of cultural dimensions such as individualism and collectivism or uncertainty avoidance [257]. Such cultural dimensions can be analysed to design components of social networking sites customised to related user traits. Considering cultural differences is vital for businesses operating on the Global market and when localisation to certain countries/cultures is preferred over standardisation. When implementing websites targeting certain cultures, the functionality and design adaptation is paramount for improving users experience as previous studies such as [226, 257] indicate. Vasalou with co-authors [254] also emphasise needs for localisation based on findings revealing cultural differences using Facebook features.

Hofstede's model was also adopted in the study by Ji with co-authors [127] investigating relation of social network sites' functionality usage, social capital gain and cultural user backgrounds among countries including the USA, Korea, and China. While these nationalities differ in their preferences towards SNS usage, they all benefit from acquisition of social capital with the SNS usage. In respect to Hofstedes cultural dimensions, survey respondents from Korea and China scored high in collectivism, which has a positive impact on content sharing. For US respondents, Masculinity had an effect on content sharing. For US and Chinese survey respondents, Masculinity also has an impact on expert search in SNS usage.

Personal research biases and a possible dependency on further unconsidered dimensions should not be underestimated [111]. This is why other investigations by researchers from eastern cultural backgrounds might find western-oriented models to be insufficient for modelling an eastern mindset [111]. Furthermore, the validity of the cultural models should be re-assessed since the cultural values and preferences might evolve over time [228]. The World Values Survey based on two generation groups' comparison indicate that overall cultures score higher on Individualism and Life Indulgence factors (Hofstede's), and lower in "Power Distance" (PD) in average while maintaining consistent country pair differences in factor values [24]. Besides, some of the cultural dimensions or factors such as defined by Hofstede are interrelated (ID and PD), which could influence the overall modeling outcomes for the cultural groups related to these factors [228].

An online survey of Finnish young business students revealed a lack of Twitter adoption, which was explained by a relatively (to other Nordic cultures) high uncertainty avoidance and low individualism [256]. Authors mentioned that Finland is located quite close to the Reactive user group in the Lewis Model of Cultures (for instance, Japanese) and requires more support from community to communicate on Twitter more actively. In [205], different values of Japanese and Finnish cultures have been contrasted in respect to management challenges in multicultural companies while referring to Lewis and Hofstede models.

Based on the information system design expertise and prototyping of collaborative research platform, Bettoni and Eggs [23] referred to the cultural behaviour traits defined in the Lewis Model. Their analysis [23] revealed the effect of "silent novice" in relation to the Linear-active user group (Germany) when the self-perceived lack of expertise might lead to lesser information sharing. Cultural component as part of e-participation framework applied in e-learning was discussed in the exploratory study by [251]. E-learning challenges adoption and study success was discussed in [10] focusing on Arab students interested in the US education. While referring to Lewis [148], Hofstede [112] and Hall's [97] research, authors inform us about potential difficulties of Arab learners and further studies needed on gender differences of e-learning process outcomes.

We can find further references to the aforementioned sociological studies in research involving various IT advances. For instance, [181] examines synchronous teamwork, physically located at one place, and virtually located at different places. They emphasise the need of "common ground" for effective collaboration of virtually collocated teams. Geography, culture, languages, contexts are important in team works. There are also differences in work organisation due to "power-distance" differences amongst cultures such as referred in the Hofstede's model. Dress-code impressions, task or people-orientation defined by cultural impressions are also important to consider for effective team collaboration [181]. Hofstede's cultural trait of "risk avoidance" was also studied in [132] in laboratory experiments with more than 500 participants from three different countries conveying their willingness to continue a risky project with lower costs and presence of a competitive product. They suggest to assign new advancing software technologies to managers with the low uncertainty-avoidance trait [132].

We noticed a relative lack of research applications of Lewis Model of Cultures, in contrast to Hofstede's model. However, several works [205] (Finnish and Japanese values comparison), [10] (e-learning challenges for Arab students), [256] (Twitter adoption by Finnish students) employ both models. In Table 4.3 we summarised the aforementioned sociological models of cultures, by Hofstede [109] and Trompenaars [241], their application fields and related research works, their arguable advantages and drawbacks.

Research Descriptions	Limitations	Advantages	Applications
Hofstede's Model of Cultures [109]			
In [109], Hofstede defined five cultural dimensions and scored persons from more than 50 countries in respect to their behavioural traits and socio-cultural attitudes	Strong assumptions towards nation-based profiling, cultural dimensions are polarised while referring to stereotypical groups, and excluding the possible biases of the IBM corporate culture possibly affecting the survey results [163]	Large sample of respondents [109], wide adoption in research	Management in multinational corporations
An application of the Hofstede's Model of Cultures in Web design [158], case studies of several websites in respect to the cultural-awareness	Subjectivity of the websites selection, small number of the analysed websites	Suggestions to the culture-aware web design considerations	Creation of websites

Table 4.3: continued on the next page

Research Descriptions	Limitations	Advantages	Applications
Exploitation of several Hofstede's cultural dimensions (aforementioned UA, ID, PD, ML cultural dimensions) for adapting websites design towards cultural profiles of their users [227]	Software evaluation experiments were involving 97 student participants from computing and information systems fields. Additionally, the research would benefit from the user usability studies for finding out the effects of cultural factors on the user satisfaction [227]	Investigation into teamwork of the software evaluation process for different cultural user groups showing dissimilar preferences towards teamwork with persons from other cultural groups. The previous works' analysis revealed large differences in the importance of the cultural dimensions in the website development. Some interesting findings on Hofstede's applicability of the ID dimension, which seems to be not significant for Chinese users in their study. Their websites development framework is based on a "cultural fingerprint" concept, that can be further extended.	Websites design process, user evaluation, and teamwork in web development.
Hofstede's Model of Cultures' "Power Distance" and "Individualism" dimensions are explored in relation with Twitter usage of mentions by persons from top 30 countries[80]	convenience sampling	Suggestions for building culture-aware communication tools and friends recommendation services	Microblogging
Analysis of cultural dimensions and related music preferences for Last.fm users from different countries [70], suggesting to exploit social media for user preferences elicitation useful in cold-start	a danger to be confined towards lower diversity spans in case of countries such as highly scored in the Indulgence index	possibility to use only country information for recommending genres diversity in cold-start	Music genre and style preferences in Last.fm

Table 4.3: continued on the next page

Research Descriptions	Limitations	Advantages	Applications
Trompenaars [241]			
Trompenaars' [241] questionnaire involved 30000 participants from 30 companies, and helped to shed light on problem solving, conflict management, business organisation approach and rule setting across different cultures	sample included mostly managers	Large sample of respondents, discussion on ethnic differences within countries, gender and age groups	Management in transnational corporations

Table 4.3: Models of culture and their applications in management and Web research

4.4 Discussion and Conclusions

While reviewing the most prominent cultural studies by Hofstede, Trompenaars, and Lewis, we might agree on some findings on personal behavioural similarities within cultural groups, and dissimilarities between cultural groups. However, we should not be too biased to particular stereotyping assumptions. We are aware of the pitfalls of such generalisation, and cultural preferences and customs still change over time [24]. Besides, the new technological advances facilitate personal interactions across the geographical borders while helping us to learn from other cultures. Instead of focusing on cultural differences, we might focus on understanding our own cultural traits better, while being less impatient when dealing with persons coming from different backgrounds in real life communication. Understanding and learning from ourselves and others could help us to be more tolerant and improve our communication skills when dealing with people.

Even though the sociological research can provide a good starting point to improve managerial practice and business communication, we argue that adapting online web environments to the user cultural demands require more research. In this thesis, we strive to stay focused on the online communication and examine several applications of culture-awareness. We do not want to be too restricted to stereotypes describing personal behaviour. We suggest using the cultural studies as a basis for better understanding human behaviour online and further creating more user-friendly well-tailored web environments.

Next, we go into the direction of using Twitter microblogs for the discovery of user approximate whereabouts on a country-level and in respect of cultural groups, and further exploiting this information in a movie recommendation example. Since the Lewis Model of Cultures is focused on interpersonal communication, this model shows greater relationship with geographical distances (for the selected country profiles), is not heavily biased towards generalisability assumption for the national culture groups, and is relatively less explored than Hofstede's or Trompenaars', we further employ it for discussing microblogging patterns for persons from selected countries in chapter 6, investigation into microblogging communication patterns in chapter 7, privacy settings usage in Twitter in chapter 8 and building up culture-aware social recommender in chapter 9. Before then, we start with describing our methodology in chapter 5.

Part III

Contribution

Chapter 5

Methodology

> "The scientific man has above all things to strive at self-elimination in his judgements, to provide an argument which is as true for each individual mind as for his own."
>
> - Karl Pearson, The Grammar of Science [191]

In this chapter, the main research aims and objectives are provided, as well as questions to be addressed throughout the thesis. The purpose of this chapter is to describe the overall methodological approach and variables to be analysed in the next chapters, constituting the main contributions of the thesis. The tools selected are critically analysed and justified in respect of previous research works. Nevertheless, more detailed information on the experimental setup and data collection is provided in the following chapters. However, the general methodology provided in this chapter could also be adopted in related computer science and sociological studies dealing with user-generated content analysis, opinion mining, social networking analysis, anthropology and other fields, which might benefit from the vast information of microblogging data online.

5.1 Introduction

As we discussed in part I, a broad adoption of social networking platforms and the considerable amount of user-generated content allow us to study user behaviour online, gather user opinions and traits for using this knowledge in web applications for potentially improving the user experience. By applying machine learning techniques, state of the art algorithms, recommender systems, and other intelligent agents, we could provide users with personalised content and functionality when needed. Sociological factors, particularly cultural origins and related user behavioural differences are often neglected in practice while developing and using web information systems, which requires a further investigation. With the aim to close this gap, we outline the main research questions, objectives and methods to address them.

The broad research area of this thesis is social networking, which is studied from a socio-technical point of view. On the example of the Twitter microblogging system providing social networking and communication features for a global audience, we focus on the understanding how people exploit microblogging while considering different cultural backgrounds, which we refer as "user origins" throughout this thesis.

One of the main aspects is, therefore, to study connection between user cultural origins and one's microblogs usage and user preferences. We should also keep in mind, that the design of the microblogging platform also affects users' behaviour, and therefore we limit ourselves to the study of one microblogging system, namely Twitter.

Therefore, in this chapter we will focus on sociological factors which might affect microblogging users on Twitter. Based on the insights provided by the sociological study by Lewis [148][1], next we formulate our main research questions and describe methods for answering these questions.

5.2 Aims and Objectives

Previous research works on Twitter and other Social Networks demonstrates that persons from different countries exploit online services in their own manner [142, 237, 80, 4]. However, it is not clear yet how we could exploit these cultural differences for improving user experiences online. Besides, different patterns of microblogging behaviour could provide cues on user origins and thus could be used as "proxy" for

[1] we provided the reasoning behind the model selection in chapter 4 above

creating user models further exploited in web adaptation. To begin with, we further analyse a set of Twitter-specific features to determine how cultural or country origins of users impact microblogging activities on Twitter, which is our first aim or goal shown in Figure 5.1.

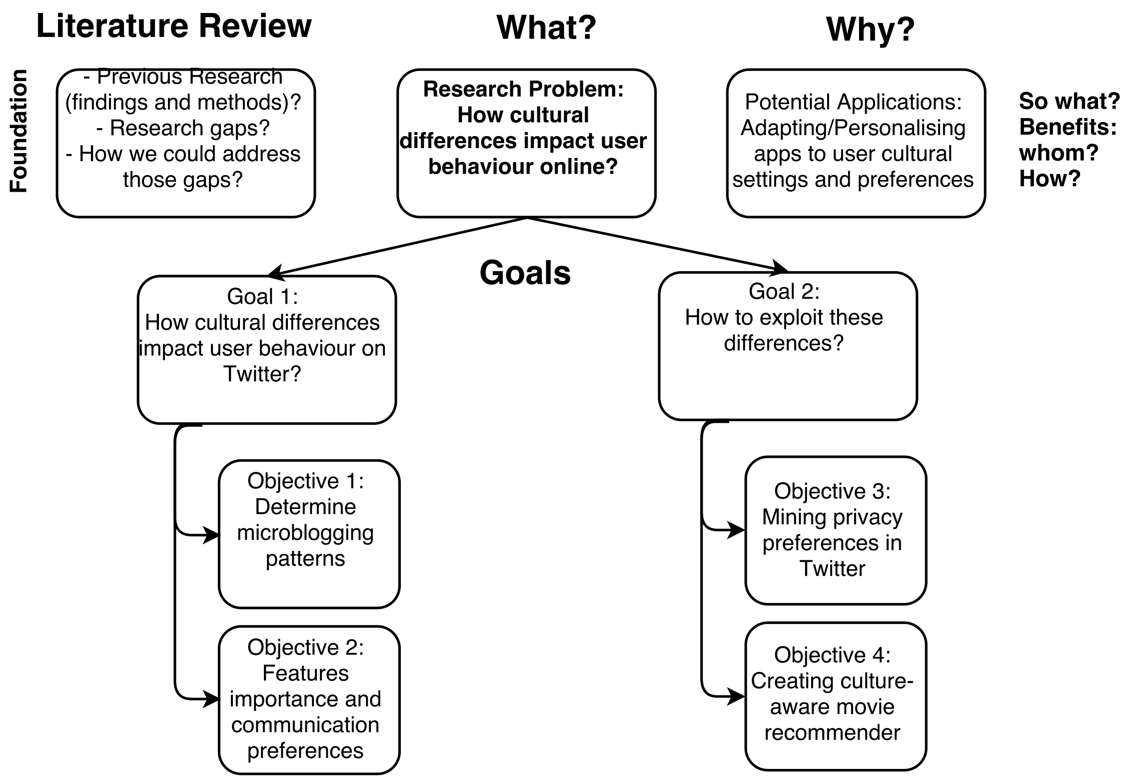


Figure 5.1: Research goals and objectives with focus on culture-related microblogging behaviour differences and preferences

These Twitter-specific features include microblogging behaviour features such as a number of hashtags or web links shared, Tweeting day of the week, content-related features extracted out of Twitter messages, and a user's social network-related features such as the number of followers and friends.

Our first objective of Goal 1 is to identify microblogging behaviour patterns, which we analyse based on aforementioned features further explained, related to user origins. With the help of classification models based on Twitter-specific features, we perform classification experiments enabling to classify users into country and cultural groups. Addressing Objective 2, the classification models created using Decision Tree and Logistic Regression techniques enable us also to analyse the relative importance of features, their impact on user origins and communication preferences prediction outcomes, respectively.

After analysis of cultural behaviour patterns on Twitter, we concentrate on the second goal, which is concerned with exploiting the detected behaviour patterns in practical applications such as user privacy settings and recommendation systems as suggested in [80]. In Objective 3 , we demonstrate how knowledge of user cultural origins can be helpful in observing user privacy preferences in microblogs. For this, we follow a set of newly created Twitter accounts for a half year period to investigate their proportion of privacy settings usage towards the end of data collection. Particularly, we analyse usage of profile protecting and geographical location sharing features in respect to inferred cultural origins.

In Objective 4 , using the Twitter search API we collect a set of IMDB movie ratings shared by Twitter users. We also infer user origins using the best model previously created, using time zone, free-text user location field, and language defined in the Twitter profile. We further explore if this information can shed light on culture-specific user preferences for movie genres. Next, we exploit this offline dataset in movie recommendation experiments. We aim to find empirical cues for applying culture-awareness practically in recommendation systems.

Objective 1: Finding Cultural Cues Online, addressed in [121]

- RQ 1.1: Could we find differences in Twitter features usage for persons microposting from different geographic regions (origins)?
- RQ 1.2: Could we exploit[2] these differences for predicting user origins on a country and geographic region level?

Objective 2: Determining microblogging features importance for predicting user origins and communication preferences [52]

- RQ 2.1: Could we exploit user contact network, i.e., friends, for predicting user cultural origins?

[2] referring to the Goal 2 since we exploit user microblogging activities for creating user origin inference models (see chapter 6 and 7) further while addressing objectives 2-4. This way we also preserve the structure of our contribution chapters in accord with the published papers.

- RQ 2.2: What microblogging features (user-related and friend-related) are the most prominent in revealing user cultural traits in Twitter?

- RQ 2.3: Could we find communication preferences in respect to user cultural origin?

Objective 3: Monitoring Twitter privacy settings usage [53]

- RQ 3.1: Are there differences in privacy settings usage by different cultural groups?

- RQ 3.2: How does protecting user accounts affect user communication styles in Twitter?

- RQ 3.3: How could user privacy preferences be exploited in real-life scenarios (discussion on security implications and related issues)?

Objective 4: Creating culture-aware social recommender

- RQ 4.1: Could we find statistically significant movie genre preferences in relation to user inferred origins?

- RQ 4.2: Could we improve movie recommendation performance when considering user origins and other item or user-related features?

The research questions we posed in this section will be addressed in the following chapters in the "Contributions" part. We aim at statistical tests and machine learning experiments for addressing the related hypothesis.

5.3 Methodological Approach

The first part of this research is empirical and based on the observations of user behaviour on Twitter. Using descriptive research methodology, we analyse microblogging features derived from user activities on Twitter. The aim resides at discovering interesting patterns of user behaviour without considering content-related features such as word distributions and topics popularity, addressed in [150, 194].

Machine learning and statistical techniques are employed to analyse microblogging behaviour for users coming from the topmost active countries in Twitter. Descriptive statistics for calculating differences in Twitter behaviour amongst cultural groups and information retrieval metrics (such as accuracy and F1-measure) for assessing results of tweet origin prediction outcomes are used. It is aimed to maximise prediction accuracy. Statistical t-tests (Welch's or Student's t-test was selected with regard to variances and distributions of samples compared) are applied for comparing user microblogging behaviour among different user groups as defined by user cultural stereotypes. Regression analysis is used to investigate the relationship between the analysed variables.

The proposed approach is also evaluated with the help of prototyping (Figure 5.2). For assessing the impact of considering user cultural backgrounds while providing recommendation results, we create user profiles considering cultural differences of the users. We compare and contrast recommendation performance for different user modeling and recommendation strategies. As a baseline, we do not include culture traits into the model and assess the simple recommendation strategies aforementioned (we call them average-based).

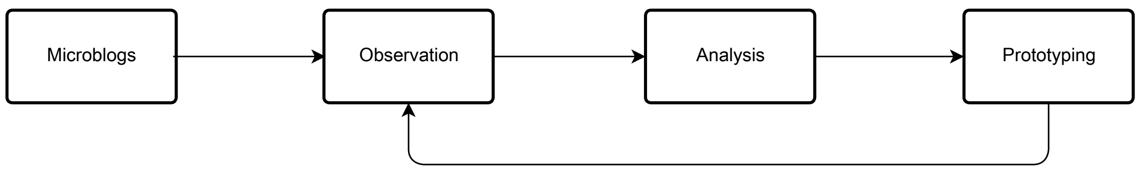


Figure 5.2: Methodology

5.3.1 The Model of Cultures

As stated in [198], given current globalisation and migration processes, it is difficult to define the term of culture. Applied to IS and Informatics research, sociological theories are explaining cultural behavioural differences, usually based on country surveys and employing cultural dimensions to stereotype individual behaviours. These theories include Hofstede's Cultural Dimensions [111] and Lewis Model of

Cultures [148] discussed in the previous chapter. Even though the limitations of creating national stereotypes and overall sociological model's usage are critically discussed [163, 273], their application are considered, for instance, in e-learning [190] and e-commerce research [56].

We employed the Lewis model of cultures [148] describing how persons belonging to different cultural backgrounds are unlike in their interpersonal behaviour. For instance, Asian cultures, such as Japan or Vietnam, are defined by Lewis in their culture dimension as Reactive (RE), since they are generally considered to be courteous, accommodating and good listeners. In addition, Lewis defines a Multi-Active (MA) and Linear-Active (LA) cultural dimension. While persons described as MA focus on interpersonal communication and are generally considered as emotional personalities, LA persons focus on working with facts and planning activities [148].

Therefore, each of the cultural groups/nations can be described with the help of "cultural dimensions". In the following, we use cultural group and dimension terms interchangeably. We thus define cultural groups based on geographical regions and particular countries in the Appendix E listing the labels assigned by human assessors with knowledge of the Lewis Model of Cultures).

5.3.2 Experimental Setup

Firstly, we aim to find microblogging behavioural differences for people publishing tweets from different cultural origins. We select Twitter users whose tweets originate from respective geographic locations provided within their tweets' the meta-data. For this, we define a list bounding boxes for geographic locations corresponding to the to the big cities in the selected countries. With Twitter location-based filter [3], we collect the tweets posted in these locations. Further, we use the collected tweets' meta-data to identify and select a set of authors as described on page 94 in chapter 6 "User Origin Prediction". Next, their tweets were collected using the Twitter Streaming Application Program Interface (API). User profiles based on the metadata and content of the tweets were created. It is important to mention that we

[3] `https://developer.twitter.com/en/docs/tutorials/filtering-tweets-by-location`

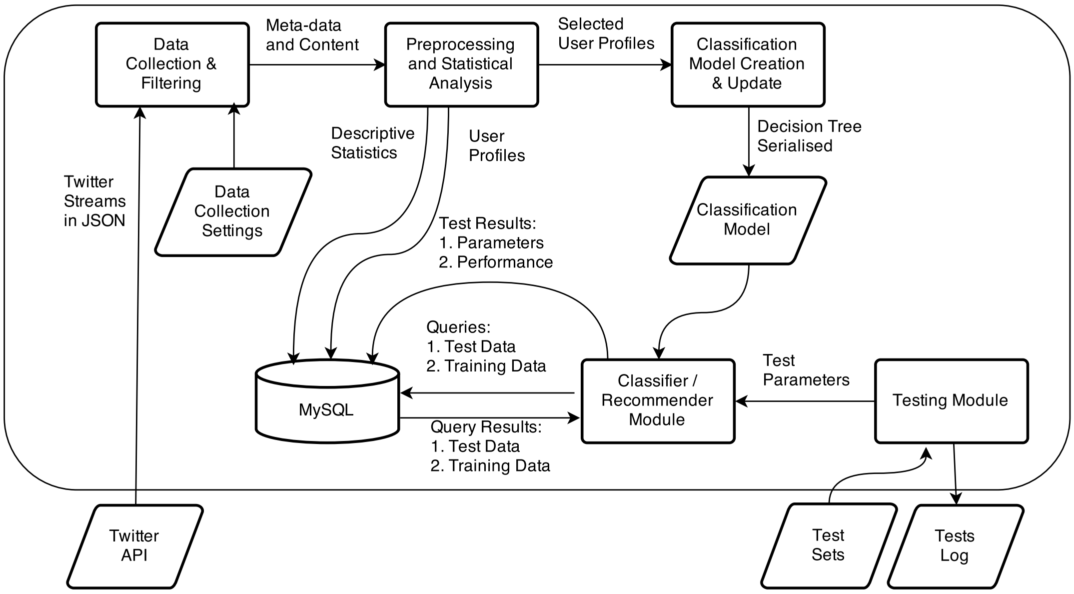


Figure 5.3: Experimental setup (simplified)

anonymise particular users' data before storing it into the database and we perform this in the Pre-processing and Analysis component shown in Figure 5.3.

To analyse how the cultural behaviour patterns differ amongst user groups, we employ statistical t-tests. Student's t-test can be used to compare two independent samples with normal distributions and equal variances, in case of variance equality assumption violation, Welch's test [260] can be performed, however, requiring that the data is normally distributed [269]. When having large datasets, the violation of the normality assumption can be ignored [86]. For testing for normality, we employed Shapiro-Wilk's test recommended as a good choice in [240, 195, 276]. D'Agostino-Pearson omnibus test considering distribution's skewness and kurtosis, and called as "moment test" in [195] and implemented in Scipy.stats Python package [215] referring to the test descriptions in [54, 55].

Next, the microblogging patterns found are used to classify users in their respective user groups including countries and cultural dimensions (regions) as defined in [148]. We assess the classification performance using information retrieval metrics such as accuracy, precision, recall, and f-measure [67] while considering the possible biases of used data samples as explained in [220] , while assessing the recommendation strategies. We compare the country prediction results with the geocoding information provided by Twitter. Additionally, we perform human assessment (see appendix H.2).

Inter-rater Reliability. To assess the accuracy of the human and automatic annotations, we employ Inter-rater reliability coefficients. Particularly, we considered Cohen Kappa [42], Fleiss' Kappa [74] and the most recent and robust metric Krippendorff α [140], which is applicable also for cases with missed data.

In 1960 Cohen introduced an inter-rater reliability measure to account for a chance agreement which is not considered when using only percent agreement measure [42]. The Cohen Kappa is applicable to compute agreement between two raters which each classify or annotate a set of items into mutually exclusive regions or categories [42]. Cohen's Kappa values range from K=0 when the agreement is achieved by chance to K=1 for perfect agreement. To decide on the agreement levels, an arbitrary benchmark scale for Kappa values was developed by Landis and Koch [145]. They advised to interpret Kappa values in the range from .41 to .60 as moderate, from .61 to .80 as substantial, and from .81 to 1.00 as almost perfect.

Fleiss' Kappa extends the inter-rater reliability assessment to more than 2 annotators and can be applied to the categorical data [74]. It is calculated as a degree of agreement achieved above chance divided by the degree of agreement which can be achieved by chance [74]. Fleiss' benchmark for Kappa values defines Kappa values below of .40 as poor agreement level, between .40 and .75 as adequate, and above .75 as excellent [73].

To account for Fleiss' and Cohen Kappa limitations while dealing with missed values, being robust on small sample sizes, any number of observers and any measurement data types, Krippendorff proposed his α coefficient [140, 141]. He argues that averaging all categories can obscure "unreliable categories" [140]. Krippendorff α takes into account fraction of observed to expected disagreement [140]. He suggests using $\alpha >= .80$ threshold for good reliability, otherwise $\alpha >= .67$ can be considered of acceptable quality [140]. Due to its underlying assumptions and paradoxes such as low α when observing high agreement percentage in practice, Krippendorff's α is also criticised in [281]. We, however, employ Krippendorff's α because of its robustness, applicability to more than three annotator cases, and working with the missed data. We, however, compare α when possible (when two raters are considered, and no missing data occurs) with Cohen and Fleiss Kappa coefficient results. Considering several metrics enable us to critically discuss attained agreement levels, which benchmarks scales are arbitrary, debatable and thus requiring more attention when interpreting results.

Evaluating Performance of Recommendation Strategies. The knowledge on user locality and cultural context is further exploited in the recommender system. The recommendation strategies are based on pre-filtering and user modeling as described in [185] while considering user origins identified. For evaluating recommendation strategies we employ metrics for assessing rating prediction error (such as Root Mean Square Error described in [235]) and other metrics discussed in chapter 9.

Hull [117] described two-sample paired comparison tests usage in evaluating information retrieval systems. He described paired t-test, the sign test, and the Wilcoxon tests while referring to the significance testing for comparing performance metrics. To compare performance metrics of different recommendation strategies, we employ two-sample paired t-test when having relatively large datasets. Besides normality assumption paired t-tests require continuos variables, independence of observations and no outliers [232]. Our experimental setup allows us to satisfy the underlying assumptions of the paired t-tests except normality. When having large sample sizes, we can employ t-tests even with the not normal distributions as discussed in [86]. Hull [117] referred to t-test as generally more powerful and robust to violations of its normality assumption. Having more than 200 users in our offline user tests, and 19 cases for the timeline tests, we performed paired t-tests and Welsh t-tests respectively.

Overall, the experimental setup (Figure 5.3) includes general data-mining steps such as exploration (data selection, filtering, pre-processing and analysis), model creation and evaluation (using Testing Module), and application of the model for classifying the users into their respective cultural profiles or recommending related contents or functionality. The prototype works with the classification and recommendation models, which parameters are guided by the set of tests. The recommendation process is envisioned to exploit the user origin predictive (classification) model for extracting user context including information on user country of origin and cultural dimension.

Selected Locations. For the users' selection we adapted the process described in chapter 6 "User Origin Prediction". We selected five countries also included in our top-10 countries in our sample dataset (Appendix F) and located at apexes or close by as defined in the Lewis Model of cultures [148]. For instance, we selected five countries, including Japan, Germany, the USA, Brazil and Spain, which we associate with the cultural dimensions of Lewis' model (Figure 6.1).

To find Twitter users belonging to the selected countries, we employed Twitter Streaming API providing samples from public data streams. After data collection, we selected users with a defined number of friends, followers, tweets and the location field mentioned the corresponding country in their Twitter profiles. We also focused on geographic locations to include large cities for the defined countries. In chapter 6 we further describe our data collection and the pre-processing process we employ for Twitter features usage analysis based on the dataset "Features" [4].

Top Countries in Twitter. In chapter 8 "Privacy Settings Usage in Twitter" and chapter 7 "Communication Preferences" we focus on the top most active countries in Twitter. To select users, we listened to the Twitter sample stream for a particular time period and saved user identifiers with the defined geographic locations in their tweets. Since it is important to understand temporal patterns of possible changes in user behaviour, a long-time study is further performed based on the Twitter data for about half of year. This allowed us to compare how the behaviour of particular users and user groups changes over time. Particularly, we were able to follow changes to users' privacy settings to find out user culture-specific preferences towards protecting their Twitter profiles (chapter 8).

[4] All data sources are listed in Appendix D

Twitter Access Details.

Number of Get Requests per Hour	Authentication Required?
REST API	
150	Unauthenticated
350	Authenticated
Search API	
Search limitations are imposed on IP addresses.	
Streaming API	
– (*)	Authenticated

Table 5.1: Sample access methods as defined in [246]

Twitter provides an Application Program Interface (API) to its services having specific rate limitation outlined in the Table 5.1 and usage purposes. Representational State Transfer (REST) API could be used in web applications for posting tweets and following other users and amongst other, online activities [246]. REST API could be used in two modes, authenticated and unauthenticated, having different rate limitations for retrieving information from Twitter. Tweeting and favouring activities do not contribute to rate limitations. When hitting the REST limit, some of the features would not be available for approximately one hour [246]. Search API provides search functionality, which is also accessible via the web interface and preferred for usage.

Streaming API is used for access to the real-time content of the Twitter sample [246]. In accord with [250], there are three end-points to access Twitter samples including "filter" providing access by defining particular users, geographic boundaries and keywords, "sample" providing a random set of about 1% of all tweets, "firehose" providing access to all public tweets and with special permission. A complete access to Twitter data could be provided by Gnip [5]. Streaming API does not impose rate limitations on sample tweets provision. However, there is a limitation (*) for developers, which should not create too many non-persistent connections using the same account settings while debugging their code as suggested in [243].

[5] https://gnip.com/sources/twitter/

Twitter API could be accessed with the help of developer libraries available for the number of programming languages including quite popular web development languages such as PHP, Python and Java [247]. Twitter API returns a text string in the human-readable JavaScript Object Notation (JSON) format, which could also be parsed in programs. However, Twitter rules state that it is forbidden to export Twitter content into web service datastore or cloud-based web service solutions [244]. Also, Twitter content datasets could made available for download when provided with user and tweet IDs [244].

Data Collection Methods. For collecting Users Dataset, we listened to Twitter Sample Stream using Streaming API. This gave us a sampling of public tweets in real time. We consume all the received tweets, store them for further analysis into the MySQL database while performing anonymisation of usernames and descriptions when applicable. Additionally, the streaming API's filter follow option allows to follow up to 5,000 users. This way, we were able to follow up to 20 thousand users while employing four personal Twitter accounts. Twitter REST API was employed to retrieve information from the particular user profiles. We, however, observed to stay within Twitter-specific rate limitations to ensure access to Twitter data.

Overall, we collected four datasets outlined in Table D in Appendices. The first dataset "Features" (referred in chapter chapter 6 "User Origin Prediction") consists of aggregated user profiles with statistics on Twitter-related features usage for users tweeting from Brazil, Japan, USA, Spain and Germany in 2012. "Communication" dataset (chapter 7 "Communication Preferences") was collected in 2014 to study user communication preferences. The "Privacy" dataset (chapter 8 "Privacy Settings Usage in Twitter") was created during half year of data collection in 2016 when user profiles were visited on a daily basis to monitor privacy settings usage. The "Recommender" dataset included IMDB movie ratings for building culture-ware recommendation strategies and also offline test results of the recommender system (chapter 9 "Culture-aware Social Recommenders").

User Profiles and Features Selection. Big data available online can provide web developers with user-generated content and past online user behaviour in a case when user profiles are openly available for analysis. All these data also enable to derive user opinions [172], geographic locations [13], topics of interest [272] amongst other elements of interest for user profiling. Considering Twitter microblogs, openly available user profiles enable web applications to gather the following information [6]:

- Tweets' content (derived from status messages);
- User activities such as sharing of hashtags and URLs;
- Precise Geographic location and free-form location information (when shared);
- Time of postings;
- Used device and application information;
- Metadata on time zone (*);
- Preferred language (*);
- Social contacts network (friends and followers);
- User name (*) and description (*)

User microblogging activities and thus one preferences can be mined via analysis of the usage of Twitter-specific features, which in this project also comprise a user profile attributes (their selection justification is presented in chapter 6 "User Origin Prediction"), such as follows:

- Hashtags to analyse how often users share hashtags and organise their content (derived from status messages);
- URLs to analyse links sharing behaviour;
- Languages to analyse foreign language usage;

[6] This is not an extensive list of the all available information, more meta-data elements are available with a user profile and published tweets. Some data herein marked with asterisk *, is also available while accessing the protected profiles as for April 2018 tested using Tweepy python library version 3.3.0.

- Geo-location to analyse users mobility;
- Time of posting, days of week;
- Friends;
- Followers;
- User mentions;
- Replies to other users;
- Retweets of other posts.

Classification into Countries of Origin. To approach selection of a machine learning algorithm, we follow the process depicted in Figure 5.4 and explained in [64]. Having a well-defined categorical target variable of country or culture group, we selected a supervised learning approach. To classify users into their respective groups, Decision Trees DT technique is employed taking in user profiles for generating training and testing datasets. In our preliminary tests shown in Figure C.1 (appendices), DT showed the best accuracy and good computation time when predicting user country groups when using the count-based features (chapter 6 "User Origin Prediction") such as a number of hashtags or user mentions' shared. When dealing with protected user profiles (privacy settings analysis in chapter 8 "Privacy Settings Usage in Twitter") or increasing the number of countries predicted (communication analysis in chapter 7 "Communication Preferences"), we assessed text-based classification approaches using features such as tweets' content and free-text location field.

Culture-aware Recommendations. The previous research works suggested that users from different regions have their preferences towards technology usage. Microblogs usage differences are discussed in [193], and communication flows between countries is analysed in [79] suggesting to exploit knowledge on user cultural traits for building recommendation systems. The exploitation of the cultural-specific user traits for building culture-aware software applications might provide advantages

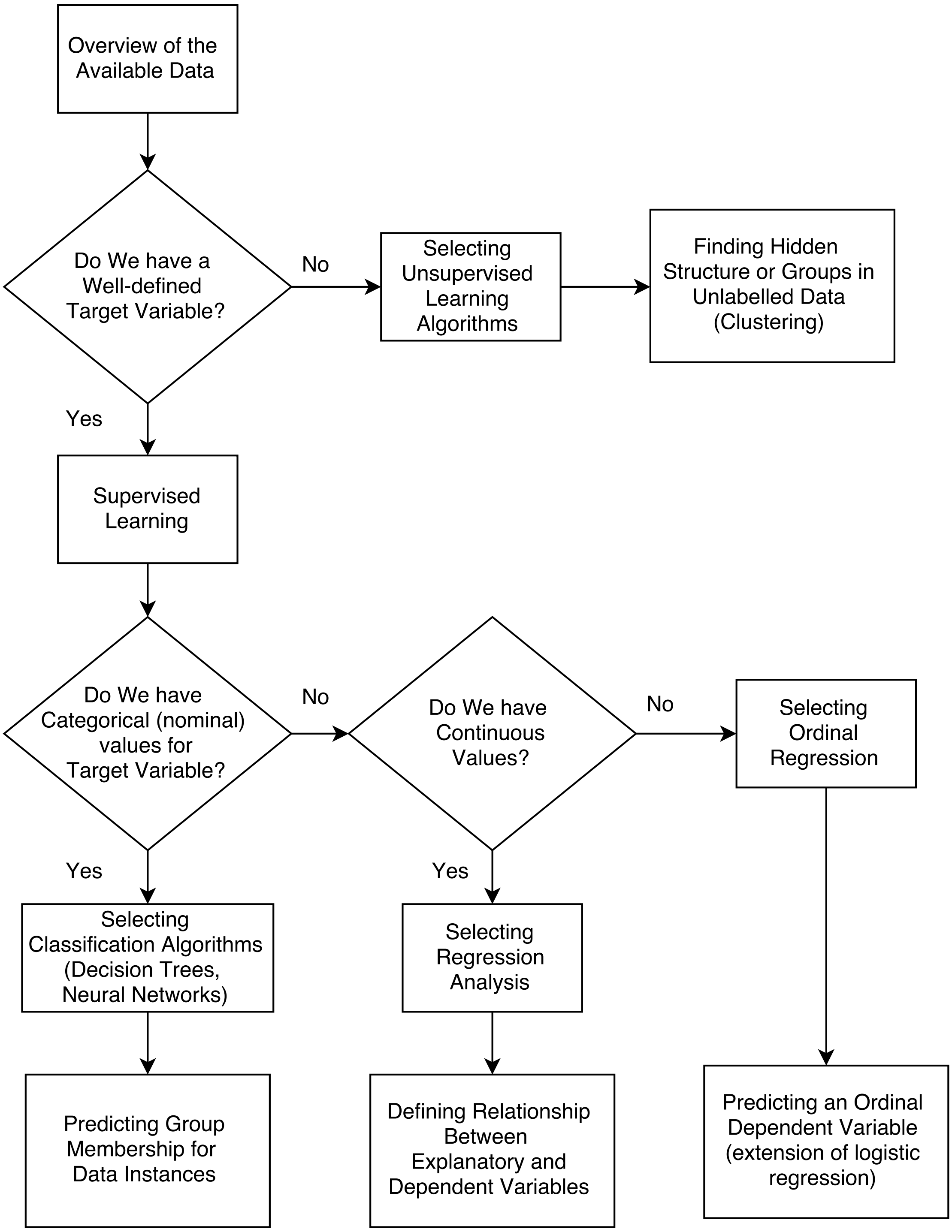


Figure 5.4: Selecting machine learning approach

when the user requires relevant content or functionality, which relate to one's own country or region. In this respect, we aim to exploit knowledge on user cultural user origins such as extracted out of user microblogs for building up culture-aware recommendation strategies. Possible practical applications include friend and Twitter content. We experiment with movie recommendations using movie rankings extracted out of Twitter messages similarly to [60]. Additionally, we assign inferred country locations and respective cultural regions. In chapter 9, we realise several recommendation strategies considering the inferred localities and using regression models for predicting movie ratings.

5.4 Discussion

On the question on ground truth, how do we define the user origin and ground truth data while training our country predicting models? We exploit country locations for geographically tagged tweets as ground truth labels. This is the class parameter exploited by the supervised learning algorithm, which is trained to label users into their countries of origin automatically. In respect of sociological studies, we define cultural origins of users as their nationalities. There are two major assumptions which have consequences on how we deal with them.

Our first assumption is that users tweeting from a particular country location are its nationals. I agree that that might be not true for some of the individual users, which includes travelers and immigrants microblogging from countries which are not their origins. One suggestion would be to manually label[7] a random sample and find out the fraction of users tweeting from their countries of origin. Another way would be an automatic inference when having more data from several social networks. For instance, we might consider a Facebook graph and get user origins from there. However, in the case of social networking data, how could we ensure that the user data available there is correct? Would this approach be prudent to define a user country of origin as coming for example from The Netherlands, both provided

[7]In chapter 9 "Culture-aware Social Recommenders" on page 184 we describe our human assessment process (see Table H.4 for results) using a small sample of users shared their geographic locations in Twitter meta-data.

in Twitter and Facebook? In my case, for instance, I would have NL locality in Twitter, and, my Facebook location is also The Netherlands having some Dutch connections as well. In fact, social networking data can be misleading and, even wrong. Even with human assessors, we could not be completely sure that the data is indeed the gold standard ground truth. This is why I would not suggest using these two approaches for evaluating our ground truth dataset.

However, to a certain extent, we can address the limitations of ground truth data collected from Twitter microblogs with geographically tagged tweets. For instance, in chapter chapter 7 "Communication Preferences" we established a rule to consider only tweets with their author languages matching with the top languages used in their country locations provided by Twitter [8]. Would be a model trained on geographically tagged dataset generalised to other instances without geographic locations attached? The generalisability of training on geographically enabled tweets was discussed by Han with co-authors in [100] showing that inclusion of non-geographically enabled tweets could even further improve the performance.

A further assumption could be, that when we think about the big data we are dealing with, we might reason that the majority of users microblogging from a particular country is indeed its nationals. This is my personal preference. This way we still do not have the highest standard ground truth, but we consider that our assumption would result in some classification error, which we do not disregard but further address. While increasing the number of users, we try to minimise the effect of the incorporated error. Related online user studies when found would be particularly useful. However, in the absence of such data to get some confidence in our estimations, a human assessment [9] on a small sample of users having geo-tagged content can be performed.

To recap, we do not have gold standard ground truth since we cannot guarantee that social networking data is truthful and reliable as such. Undoubtedly, not all our users originate from countries they tweet. However, big data could bring us closer to the truth, not golden standard, but an approximation based on the law of averages.

[8] As we further observed in Table 7.2 this rule helped to achieve a better performance for almost all feature combinations analysed

[9] Please refer to chapter 9 (page 184) and Appendix H.2

One of the relevant findings by Alex with co-authors [13] reports that almost 70% of tweet locations are within 99 km distance from the locations defined in the free-text location fields in the Twitter profile, and only 14% of the distances are greater than 999 km. To evaluate the success of the country or culture level classification models trained on a less than optimal quality but big data, I would suggest using a test set consisting of control groups' instances in which several assessors agree on user origins. In this case, we might employ solutions such as Amazon Mechanical Turk and individual human assessors.

5.5 Conclusion

Above, we described our research methodology to uncover and employ culture-specific microblogging behaviour patterns in Twitter. We proposed to exploit data mining techniques and statistical analysis to model and compare microblogging behaviour for persons from defined geographic regions. We use Twitter API to extract Twitter-specific features, which we pre-process and anonymise to create user profiles while protecting user privacy. The aggregated user profiles we further used to classify users into their respective country and culture groups (chapter 6) in order to further analyse user communication preferences (chapter 7), privacy requirements (chapter 8) and culture-aware recommendation scenarios (chapter 9).

Chapter 6

User Origin Prediction

> "The communications and information revolution is a dream come true for data-oriented cultures."
>
> - R.Lewis, When cultures collide [148]

This chapter is based on my publication "A User Modeling Oriented Analysis of Cultural Backgrounds in Microblogging" [119] (best paper award), presented at ASE Social Informatics in Washington D.C. in 2012. Next, I further extend the publication's contents with new details on data analysis and results.

Adaptive applications rely on the knowledge of their users, with their needs and differences. For instance, the training processes can be adapted to user origins using information on their cultural background. Our goal is to gather culture-specific information from publicly available microblogging content. For this, we analyse culture-specific microblogging behaviour patterns. We monitor the usage of Twitter-specific elements including hashtags, web links and user mentions. We analyse how users from different cultural groups employ these elements when they tweet. On the analysed user groups from different regions, we identify distinctive microblogging patterns. Our findings reveal a culture-specific user behaviour on Twitter which we explain regarding previous sociological research. Since our results enable us to distinguish between different cultural origins of user groups; we propose a culture-oriented user modeling approach which enables us to capture user microblogging behaviour patterns. User microblogging behaviour provides an outlook into user preferences towards sharing content with others, time preferences and social networking attitudes.

6.1 Introduction

Adaptive applications such as e-learning environments benefit from the knowledge of the cultural backgrounds of users. For instance, e-learning applications aiming to work with students from different cultural backgrounds benefit from a representation of culture-related aspects of the users. One of the case-studies of the ImReal[1] project involves learning how to effectively communicate with people from other cultural backgrounds. In this case, culture-oriented user modeling could take place by considering cultural aspects of users and using them in adapting the application behaviour according to the user needs. However, cultural-oriented user modeling is not a trivial task, since it requires an in-depth understanding of user characteristics in relation to the concept of culture including nationality, religious and political views, education level, country of living and other residence locations which influence the real-life user experience [197].

As result of user modeling, user profiles representing user characteristics are created and used to adapt applications to user needs. In case user-related information cannot be retrieved directly from the user or is not available, adaptive applications might exploit user data derived from external sources like social networks and microblogs. For instance, Twitter content can be used to create user profiles describing user interests [2]. Twitter profile data can provide information on a user's geographic locations and use of languages. Related data may also be extracted from microblogging content published by the user.

User preferences according to user's location can be extracted from microblogging content and this information stored in the user profiles. However, would it also be possible to derive culture-specific behavioural traits based on user microblogging activities? Can we ascertain culture-oriented behavioural patterns of user behaviour on microblogs? Does the information derived from Twitter allow us to differenti-

[1] `http://www.imreal-project.eu/`

ate users belonging to different cultural groups? These questions motivated us to investigate how to mine cultural patterns of user behaviour on Twitter.

For this, we analyse microblogging behavioural patterns and relate our findings with sociology research on cross-cultural communication [148]. We assume that communication differences could be reflected in the way people blog. For this, we create stereotypical cultural background models reflecting their behavioural patterns on Twitter, based on a set of users with defined geographic locations.

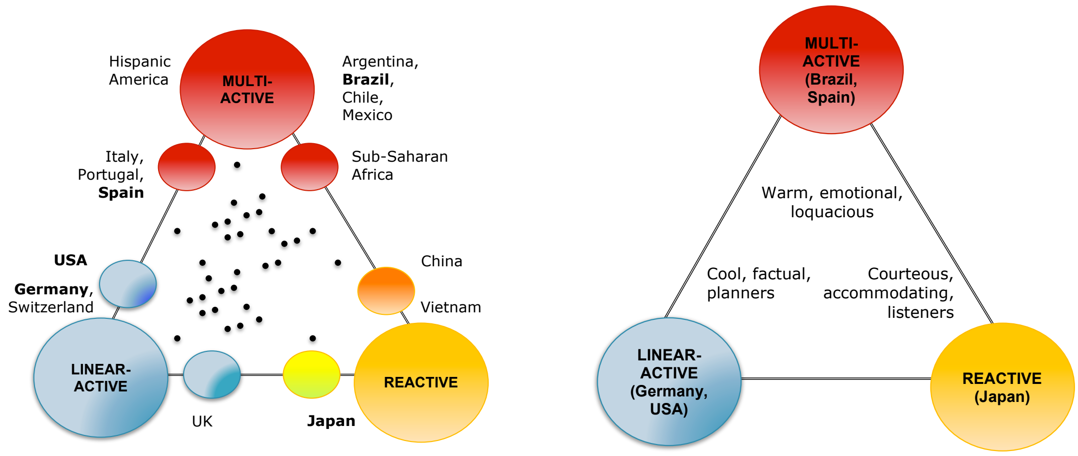
These stereotypical models allow us to get insights on user microblogging behaviour and its differences among cultural groups. Our main contributions discussed in this chapter include the following:

- An analysis of user behaviour on Twitter for five user groups of different cultural origin.
- Culture-specific microblogging patterns as explained by culture-related characteristics from sociology research by Lewis [148].
- A Culture-oriented User Modeling approach based on user behaviour in microblogs and its experimental assessment.

Next, we briefly outline the scope of the study and related work. Then, we describe our research methodology and the experimental setup for our culture-oriented user modeling experiments. This is followed by an analysis of Twitter features usage for selected countries and user groups, and report on the quality of the created models, which is based on experiments predicting users cultural origins. We conclude with our main findings on user behaviour for the five cultural groups and provide insights into further user modeling research directions considering cultural behavioural patterns.

6.2 Research Methodology

First, we outline our conceptual framework for culture-oriented user modeling, which is based on the Lewis Model of Cultures (Figure 6.1). Then, we formulate research



(a) The Lewis Model

(b) Adapted Model

Figure 6.1: Lewis Model of Cultures [2] (simplified and adopted from [148])

questions in relation to findings from previous research, and we describe the experimental setup and dataset used.

6.2.1 Conceptual Framework and Approach

In [148] Lewis analyses the cultural personality dimensions in relation to the country of origin or nationality. For instance, the multi-active Hispanic users group includes citizens from Argentina, Spain and Brazil, the linear-active group from Germany and the USA. The reactive dimension reflects the group of citizens from countries such as Vietnam and, to a lesser extent, Japan. Persons from the linear-active group share some similarities amongst each other such as a focus on planning activities, factual information, and respect towards authorities. The reactive group can be associated with polite behaviour and conflict avoidance. Multi-active persons generally tend to show their emotions and multi-task. Citizens from other countries are placed between these extreme groups and each person can be described in terms of reactivity, linear-activity and multi-activity traits [148].

The Lewis model of Culture is represented in the form of a triangle with corner points depicting the cultural dimensions mentioned above, as shown in Figure 6.1 (a). These cultural dimensions reflect differences in the way people with different cultural backgrounds communicate [148]. In our experiments, as shown in Figure 6.1 (b), we selected users from Germany and Brazil located in the apexes of the Lewis model of Cultures and representing linear-active and multi-active user groups respectively [119]. The USA and Spain were added to the respective user groups even though these countries are not located directly at the apexes of the triangle. This allows to analyse the behaviour of the aforementioned user groups and how their behaviour differs in respect of the Lewis research findings. We selected Japan for representing a reactive user group even though it is not depicted in an apex of the Lewis model, since it is listed as one of the top most active countries on Twitter as reported by Semiocast [218]. This is why we selected Twitter users from Germany and the USA for representing the linear-active group, users from Japan for representing the reactive group, and users from Brazil and Spain for representing the multi-active group, as shown in Figure 6.1 [119]. The inclusion of five countries enabled a comparison of the user groups originating from these countries. User groups from countries belonging to the same cultural profile, corresponding to the Lewis model, are further aggregated for comparison. We believe that this approach can be used for further modeling user profiles of users from different countries into the three cultural profiles according to the Lewis model. This could be advantageous for applications targeting cultural differences [3].

To acquire knowledge on user traits related to the cultural background of a user, we propose to mine them from microblogging activities of the user. Based on the analysis of microblogging behavioural patterns, culture-oriented user modeling can be performed. In result, user profiles with information on culture-specific user traits and preferences can be created and used in the adaptation process as shown in Figure 6.2.

[3] As we will further find out in chapter 9, cultural-group based pre-filtering can help in improving recommendation performance in cold-start recommendation situation when no previous user history logs yet available

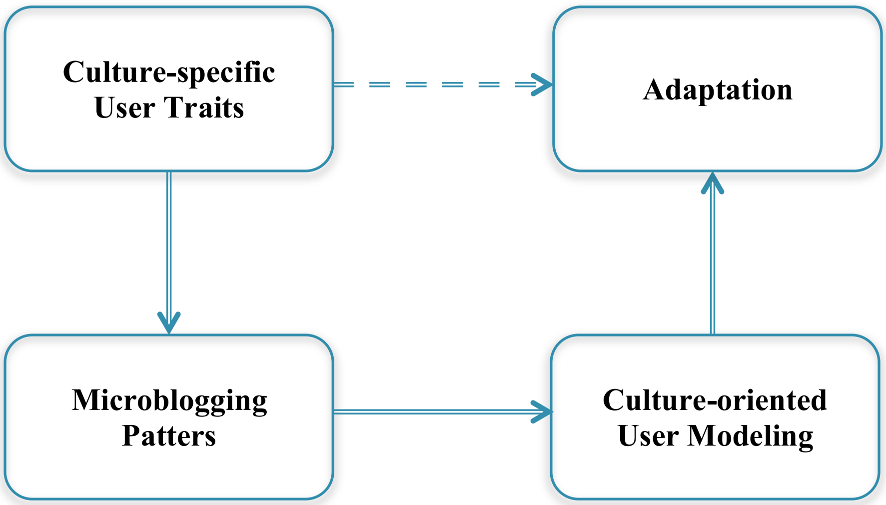


Figure 6.2: Culture-aware adaptation with automatic inference of user locations

6.2.2 Research Questions

Richard Lewis in his sociological study [148, 149] describes culture-specific behaviour patterns in interpersonal communication. We base our investigation on the idea that personality traits as defined by Lewis [148] are also reflected in the way how people blog on Twitter. The previous works [211, 254] inform us about cultural differences in music listening patterns derived from Twitter posts, usage of Facebook features and time spent differ respectively across countries. Similarly, we seek to find culture-specific styles of using Twitter features such as user preferences of sharing hash-tags or replying to others.

Sharing Content. Huang with co-authors [114] stated that hashtags could be used not only for conversational purpose, but also for organising content, in which case the hashtag standard deviation time is higher than the standard deviation time of hashtags used in conversations. Also, in [78] we read, that users from less individualistic societies might refrain from using hashtags since they do not like their tweets to appear in trending topics. Poblete with co-authors [193] stated that the USA is the first country in their list, leading in Uniform resource locator (URL) sharing. We assume that this can be explained by the linear-activity characteristics of these users. In his work [148] Lewis wrote that linear-active persons are "data-oriented" and prefer to work with factual information, while reactive users are more "dialogue-oriented". This is why we consider hashtags and URLs for comparing linear-active users with others.

Foreign Languages Usage. In one of the related microblogging research [193] authors found that the English language is the most popular language on their dataset, created from Twitter content and accounts for more then half of the tweets published by users in the ten countries analysed.

Taking into account the widespread usage of the English language on Twitter and the challenges regarding automatic language detection as stated in [193], we investigate how the number of detected languages differs between the analysed user groups. We are interested in comparing foreign language usage of multi-active and reactive persons, which in accord to Lewis model [148] are both people-oriented, while multi-active persons are more extroverted and loquacious.

We used langdetect tool which is the language detection library for Java [222] available at Google Code. Lui and Baldwin [154] reported that langdetect is implemented as a Naive Bayes classifier trained on the Wikipedia data and quite robust for short documents such as Twitter microblogs with the accuracy of 93% when evaluated on 9659 Twitter messages posted in six European languages. Fujii et al. [76] evaluated langdetect with tweets published by 109 bloggers overall using 15 languages (majority also exploited European languages, except of two Japanese and one Afrikaans language detected) and reported language detection precision of 97% and recall of 80% when automatically identifying languages with the highest probability.

Tweeting Mobility. Since tweeting mobility is also interesting to investigate as mentioned in [193], we analyse how users from different countries use Twitter while traveling. Since linear-active persons can be very conscious about effectively allocating their time [148], we assume that linear-active persons might use Twitter on their travels. We, therefore, compare their tweeting behaviour with reactive persons.

Posting Time. Tweeting time during weekends or weekdays was important to relate with the reactive user, who generally has a different perception of time, being very punctual and polite, as outlined in [148]. This is why we assume that reactive persons might employ Twitter more on weekends compared to other persons, particularly linear-active persons, since multi-active persons tend to do things "at the same time", as described in [148].

Referring to other Users. Since findings by Poblete et al. [193] show that Japanese users mention other users the least, we investigate user mentions employed by the analysed user groups. Also, Lewis [148] states that persons from Japan generally employ fewer names than persons from Western countries.

Social Network Size and Conversations. We also consider conversation and social network features for analysing user communication on Twitter. Boyd with co-authors [26] stated that retweets could be used to give credit to other bloggers or even self-promotion. A study by Poblete with co-authors [193] shows that Japanese users retweet the least. We investigate the retweet frequency of the analysed user groups and compare the results with the previous research by [193]. Besides, since multi-active and reactive persons are people-oriented as stated in [148], we compare these two user groups. Since reactive cultures value silence and more in-depth content [148], we assume that reactive persons might refrain from retweets in opposite to multi-active users. This is why we hypothesise that reactive users might reply more since they are indicated as being very good listeners in [148].

In a nutshell, our main research goal (addressed in this chapter) consists in analysing Twitter microblogging behaviour for users from different cultural groups as defined in the Lewis model. We study how users from linear-active countries (Germany and the USA), reactive countries (Japan) and multi-active countries (Spain and Brazil) use Twitter, and investigate how Twitter behaviour differs between these different cultural groups.

To address RQ1.1: Could we find differences in Twitter features usage for persons micro-posting from different geographic regions (origins)? and RQ1.2: Could we exploit these differences for predicting user origins on a country and geographic regions levels? of the thesis Objective 1 ("Finding Cultural Cues Online")[4], we expand them to get a more in-depth understanding of the Twitter features usage patterns and their possible application for user origins detection. We focus on content-based, activity-based and social network-based features. We explore the following research questions referring to the usage of aforementioned Twitter features:

Content-based characteristics

- RQ1.1.1 (Hashtags usage): How does the usage of hashtags differ between cultural groups?
- RQ1.1.2 (URLs): How often do users from different cultural groups share URLs?
- RQ1.1.3 (Languages): How do users from different cultural groups make use of foreign languages in their posts (as detected using automatic language detection with LangDetect [222])?

Activity-based characteristics

- RQ1.1.4 (Mobility): To what extent do the different groups of users publish tweets from different geographic locations?
- RQ1.1.5 (Weekends): How frequently do users post on weekends compared to weekdays? Do these temporal Twitter traits differ between different cultural groups?

Social Network-based characteristics

- RQ1.1.6 (Friends): How does the number of friends differ between the cultural groups?
- RQ1.1.7 (Followers): Is there a relation between the number of followers that a user has on Twitter and the cultural background of the user?

[4]see Methodology in chapter 5

Conversation characteristics

- RQ1.1.8 (Mentions): How often do users from different cultural groups refer to other users?
- RQ1.1.9 (Replies): How often do users from different cultural groups reply to other users?
- RQ1.1.10 (Retweets): To what extent do users from different cultural groups retweet?

The above research questions refer to different features which describe certain aspects of the users' behaviour on Twitter. In our analysis, we compare for which features the cultural groups exhibit the most respectively, the least profound differences. Following the feature analysis, we model stereotype user profiles and perform a series of experiments for predicting a user belonging to a specific stereotype profile. This helps us to address RQ1.2 and investigate how well our model works for different cultural user stereotypes and how can we describe user activities on Twitter in relation to the Lewis' model. We aim to achieve a reasonably high prediction accuracy while addressing the last empirical question of this chapter:

- RQ1.2.1: Could we achieve a CV accuracy above 90% while predicting user countries and cultural regions?

6.2.3 Experimental Setup

To perform culture-oriented user modeling on Twitter, first of all, we identified differences in microblogging behaviour of people from different countries. For this, we selected Twitter users who indicated their location in Twitter profiles. It is important to mention, however, that we do not profile users into gender and age groups. We analyse instead all users with a country mention defined in their profile. In [160], we read that cultural statistics on personality traits for 26 countries showed similar personality levels for men and women. Additionally, age was profiled in a similar way across countries. Therefore, we assume that users of different age and gender groups can be combined to profile aggregate cultural groups.

To find Twitter users belonging to the selected countries, we employed Twitter Streaming API[5] providing samples from public data streams. We selected users having more than ten friends and tweets, and having less than 5000 followers. For all selected users the location field mentioned the corresponding country. We also define geographic locations to include large cities such as Berlin, Hamburg and Munich for Germany, Tokyo, Yokohama and Osaka for Japan, New York, San Francisco and Washington D.C. for the USA, São Paulo and Rio de Janeiro for Brazil. For the tweets collected for the defined user groups, we analyse the use of Twitter-specific features.

Country	Number of Users	Users Posted at least 100 Tweets
Japan	4885	2984
Spain	4906	3119
Brazil	4910	2935
USA	1714	1316
Germany	2823	1644

Table 6.1: Users tweeting from five country locations (their accounts were followed from 2012-03-26 to 2012-06-01, the summary of "Features" dataset is in Table D.2)

We aimed to find behavioural patterns for these cultural user groups and explain them in relation to the Lewis' model of cultures. Overall, we performed the following steps:

- STEP 1: For the defined culture groups Germany, Japan, Spain, Brazil and the USA we selected a set of users tweeting from their respective geographic locations[6] (using longitude and latitude coordinates that define the place area).
- STEP 2: Using User Twitter Streams, for two months we collected tweets published by the users selected in STEP 1. To get a solid understanding of users'

[5] `https://dev.twitter.com/docs/streaming-apis/streams/public`
[6] `https://dev.twitter.com/docs/streaming-apis/parameters#locations`

behavioural characteristics on Twitter, we limited our dataset to users who published at least 100 tweets (performed after data collection). Our threshold of 100 tweets enabled us to aggregate user microblogging behaviour for 11998 users. This allowed us to analyse the user behaviour for more than 1000 of users for each country (see Table 6.1). In addition, during the two months crawling period, we analysed users mobility defined as tweeting from different geographic locations. We identified the country name using the Geonames API[7] and Google Geocoding API[8].

- STEP 3: After completing the crawling process, we summarised, based on 100 randomly selected tweets (this is an arbitrary number, and a further study might clarify how many tweets would be required to build Twitter activity-based user profiles) published by each user, the tweeting behaviour of each user in a user profile including Twitter-specific characteristics such as the use of hashtags, user mentions, and link sharing. On the user profiles created, we analysed with descriptive statistics how the Twitter-specific behaviour differs between culture groups. We employed t-tests for identifying which user groups behave differently and the level of significance while addressing the thesis research question RQ1.1 "Could we find differences in Twitter features usage for persons micro-posting from different geographic regions (origins)?".

- STEP 4: Next, we created classification tree models based on the features set analysed on the previous step. For this, we used a set of user profiles created in STEP 3. The classification experiments allowed us to assess the predictive value of the analysed features. We used our set of features as a set of numeric variables for building the decision tree classifiers.

- STEP 5: Finally, we evaluated the classification tree using a resubstitution method and ten-fold cross-validation. This allowed us to assess the classification accuracy and quality of generated culture-aware user profiles while addressing the thesis research question RQ1.2 "Could we exploit these differ-

[7]http://www.geonames.org/export/web-services.html
[8]https://developers.google.com/maps/documentation/geocoding/

ences for predicting user origins on a country and geographic region level?"

6.3 Analysis of Behavioural Twitter Features

A one-way ANOVA was conducted to compare the effect of the country of tweets' origin on the selected features describing Twitter usage statistics. The feature values were not distributed normally, in accord with Shapiro-Wilk test with the significance of $p < 0.05$. However, the large sample size allowed us to proceed due to the ANOVA's robustness to the non-normality for the large sample sizes [68]. The assumption of the variances equality was violated in accord with Levene's test with the significance of $p < 0.05$ for all features analysed. We performed features transformation to their square root. As the result of the transformation, the maximum to minimum variance ratios did not exceed 3 when considering the violation of the variance inequality assumption discussed in [58]. Before performing features transformation, we observed that features with the maximum to minimum variance ratio exceeding 3 among the groups included languages (ratio of 7.3) and hashtags (ratio of 3.8). An analysis of variance showed that effect of tweets' country origin was significant for all features analysed with $p < 0.01$ as shown in Table A.1. The results agreed with the Kruskal-Wallis H test results shown in Table A.2. The Kruskal-Wallis test is rank-based non-parametric test, discussed in [37], showing a significant difference of Twitter features usage among the country groups estimated with 1% significance level ($p < 0.01$).

These user features were derived from the Twitter profile of the users from the chosen country groups. Assuming that users from Japan belong to the reactive user group, the USA and Germany belong to the linear-active user group, and Brazil and Spain belong to the multi-active user group, we created user profiles based on the data collected from the user content. For establishing our hypothesis, we assumed that user behaviour on Twitter reflects the user's cultural background. For instance, tweeting time during weekends or weekdays was important to relate with the reactive user, generally having a different perception of time as explained in [148]. In addition, we also considered conversation and social network features for analysing user communication on Twitter.

Table 6.2 summarises average statistics on the analysed features and their relation with the research questions and hypothesis on a cultural dimension level (the mean comparisons on country level shown in Table A.3 in the appendices). It reports also two-sample Welchs t-test[9] (t-values) for equal means assuming unequal variances (variances equality tested based on the Levene's test with significance of $p < 0.05$) results. Table D.2 shows that all feature values were continuous (of ratio levels, having a true zero when features were not present in the user profiles, for instance, when user did not have followers, or did not share any hashtags yet; the number of friends had, however, a minimum of 10 for all our users due to our experimental setup requirement) and not normally distributed (based on Shapiro-Wilk test with significance of $p < 0.05$).

6.3.1 Results of Features Analysis

For feature usage comparison, we created user profiles (using contents and meta-data of the randomly selected 100 tweets) for each of our users assigned to their respective country and culture groups as follows:

- For defining Hashtags, URLs and User Mentions usage, we calculate the number of respective elements shared by the user overall (we compute all occurrences of the element, which might be multiple per tweet);
- Foreign languages we define as the number of languages (detected automatically) minus one (assuming that one language is the native language of the user);
- Mobility is the number of countries (when available and defined with Twitter meta-data) minus one (assuming that one country is in the user home location);
- Weekends denotes the number of tweets (out of 100 tweets) posted during the weekends;
- Retweets, Replies is the number of retweeted messages and replies respectively;

[9]Using MATLAB two-sample t-test (ttest2 with 'unequal' option for 'Vartype' parameter)

- Followers and Friends is the number of respective connections in the user network.

Next, we use the user profiles with defined feature values to compare their usage statistics across the user groups.

For this, we performed 2-sample Welchs t-test [10] assuming non-equal variances. Table 6.2 shows Hypothesis and t-test results for the feature categories comprising the Content-, Activity-, Social Network- and Conversation-based categories. The results provide t statistic values, df values for associated degrees-of-freedom, values μ_1 and μ_2 representing mean values (for each of our country/culture groups, we defined a mean value as a sum of feature values, defined in the user profiles for all users, divided by the count of users in the group) for the compared user groups on the culture-level. On the country-level, t-test results are shown in Figure A.3, where countries are denoted by their two-letter ISO 3166-1 country codes.

RQ1.1.1: Hashtags Usage. T-test results show that mean values for linear-active user groups are significantly higher than means of users from other groups. This supports our hypothesis $H_{1(a)}$ that, on average, linear-active users use hashtags the most compared to other user groups.

User Group Germany has a higher mean of hashtags usage compared with the USA user group ($\mu_1 = 34.4$, $\mu_2 = 28.7$, $p < 0.001$). It is important to mention that the means of hashtag usage are close for the user groups of the USA and Spain ($\mu_1 = 28.7$, $\mu_2 = 29.6$, $p > 0.05$), sharing more hashtags compared with users from the USA. Our experiments support the hypothesis $H_{1(b)}$ stating that reactive users use the least of hashtags compared to other user groups. The results of t-tests support the acceptance of the hypothesis, that users from Japan employ hashtags the least, on average at the very high significance level ($\mu_1 = 7.6$, $\mu_2 = 25.6$, $p < 0.001$). The country-level tests reveal that the user group from Japan shares fewer hashtags compared to the other four countries, on average.

[10] using ttest2 in Matlab package

Research Questions and Hypothesis	t	df	μ_1	μ_2	Result
Content-based: Hashtags, URLs, Languages Detected (RQ1.1.1 to R1.1.Q3)					
$H_{1(a)}$ Linear-active users share Hashtags more than Multi-Active and Reactive users	21.8	4188.3	31.9	17.4	$\mu_1 > \mu_2$
$H_{1(b)}$ Reactive users share Hashtags less than Linear-Active and Multi-Active users	-41.6	10379	7.6	25.6	$\mu_1 < \mu_2$
$H_{2(a)}$ Linear-active users share URLs more than Multi-Active and Reactive users	14.4	5109	39.6	31.6	$\mu_1 > \mu_2$
$H_{2(b)}$ Reactive users share URLs less than Linear-Active and Multi-Active users	-3.7	6471.4	32.1	34.0	$\mu_1 < \mu_2$
$H_{3(a)}$ Multi-active users employ the more foreign languages than Linear-Active and Reactive users	51.4	11145	1.1	0.4	$\mu_1 > \mu_2$
$H_{3(b)}$ Reactive users employ less of foreign languages than Linear-Active and Multi-Active users	-9.8	6044.7	0.16	0.8	$\mu_1 < \mu_2$
Activity-based features: Mobility and Weekends (RQ1.1.4 to RQ1.1.5)					
$H_{4(a)}$ Reactive users tweet less from different locations than Linear-Active and Multi-Active users	-30.3	5791.3	0.6	0.9	$\mu_1 < \mu_2$
$H_{4(b)}$ Linear-active users tweet more from diff. locations than than Reactive and Multi-Active users	15.4	4703.3	0.9	0.8	$\mu_1 > \mu_2$
$H_{5(a)}$ Reactive users tweet more on weekends than Linear-Active and Multi-Active users	22.2	6109.4	28.6	24.3	$\mu_1 > \mu_2$
$H_{5(b)}$ Linear-active users tweet more during weekdays than Reactive and Multi-Active users	-6.1	5395.3	24.5	25.7	$\mu_1 < \mu_2$
Social Network-based features: Friends and Followers (RQ1.1.6 to RQ1.1.7)					
$H_{6(a)}$ Multi-active users have greater number of friends than Linear-Active and Reactive users	-6.1	12315	310.2	355.2	$\mu_1 < \mu_2$
$H_{6(b)}$ Linear-active users have smaller number of friends than Reactive and Multi-Active users	6.2	4836.1	375.1	319.6	$\mu_1 > \mu_2$
$H_{7(a)}$ Multi-active users have greater number of followers than Linear-Active and Reactive users	-6.4	12853	315.1	376.8	$\mu_1 < \mu_2$
$H_{7(b)}$ Linear-active users have smaller number of followers than Reactive and Multi-Active users	9.7	4234.6	442.7	316.2	$\mu_1 > \mu_2$
Conversation-based: User Mentions, Replies and Retweets (RQ1.1.8 to RQ1.1.10)					
$H_{8(a)}$ Reactive users mention other users less than Linear-Active and Multi-Active users	-40.3	8052.8	46.5	71.0	$\mu_1 < \mu_2$
$H_{8(b)}$ Multi-active users mention other users more than Linear-Active and Reactive users	22.6	13037	71.6	57.5	$\mu_1 > \mu_2$
$H_{9(a)}$ Reactive users have the use more replies than Linear-Active and Multi-Active users	3.6	5456.8	27.2	25.8	$\mu_1 > \mu_2$
$H_{9(b)}$ Multi-active users use less replies than Linear-Active and Reactive users	-7.5	12837	24.9	27.4	$\mu_1 < \mu_2$
$H_{10(a)}$ Reactive users retweet less than Linear-Active and Multi-Active users	-37.8	7889.5	8.2	17.7	$\mu_1 < \mu_2$
$H_{10(b)}$ Multi-active users retweet more than Linear-Active and Reactive users	30.2	12802	18.9	11.4	$\mu_1 > \mu_2$

Table 6.2: Research questions and hypothesis test results for comparing cultural user groups (with significance level $p < 0.001$, df is degrees of freedom, μ_1 is mean value of the compared group, μ_2 is mean value of the rest of users, which could be seen as an average for testing our null hypothesis). Therefore, our tests specify that the feature statistics is either greater then or less then the value specified by μ_2

RQ1.1.2: URLs Usage. The results of the tests support the hypothesis $H_{2(a)}$. Linear-active users use URLs the most compared to other user groups. Country-level statistics reveal that users from the USA ($\mu_1 = 42.5$) employ more URLs compared to users from Germany ($\mu_2 = 37.5$, $p < 0.001$), on average.

Our tests support hypothesis $H_{2(b)}$ stating that reactive users from Japan share the least of URLs ($\mu_1 = 32.1$), on average. However, country-level statistics for users from Spain (multi-active) indicate that they share less URLs compared to users from Japan (reactive) ($\mu_1 = 30.8$, $\mu_2 = 32.1$, $p < 0.05$). Tests show a similar hashtag usage for Japan and Brazil users ($\mu_1 = 32.1$, $\mu_2 = 32.1$, $p > 0.05$).

RQ1.1.3: Foreign Languages. The hypothesis $H_{3(a)}$ is supported by our experiments, indicating that multi-active users employ the most of foreign languages automatically detected from the user-generated content compared to other user groups. The hypothesis $H_{3(b)}$ is also supported since our experiments show that reactive users from Japan employ the least of foreign languages in their tweets ($\mu = 0.16$) compared to other users. On the country-level, Japanese users employ less foreign languages followed by the USA, Germany, Spain and Brazil. Users from Brazil employ the most of foreign languages.

RQ1.1.4: Mobility. The hypothesis $H_{4(a)}$ can be accepted, since our statistic shows reactive users on average tweet less from different geographic locations compared to other user groups ($\mu_1 = 0.6$).

The hypothesis $H_{4(b)}$ can also be supported, since linear-active users (USA: $\mu_1 = 0.9$, Germany: $\mu_2 = 1$) tweet the most from different geographic locations, on average. It is important to note that all other country-level user groups except Brazil have smaller mean values for the mobility feature compared to the linear-active group. Brazil and USA user group means do not differ significantly in our tests ($\mu_1 = 0.9$, $\mu_2 = 0.9$, $p > 0.05$).

RQ1.1.5: Weekends. Statistics of the tweets fraction published on weekends demonstrate that the hypothesis $H_{5(a)}$ is supported at the very high significance level.

Reactive users from Japan have a larger amount of tweets on weekends ($\mu_1 = 28.6$) compared to other user groups, on average.

Our tests indicate that the hypothesis $H_{5(b)}$ can also be accepted since linear-active users have a smaller fraction of tweets on weekends compared to other users on average. The same trend holds on country-level statistic indicating that users from Germany ($\mu_1 = 25.3$) and the USA ($\mu_2 = 23.5$) tweet less on weekends than others on average. Interestingly, mean values for Spain and Brazil ($\mu_1 = 24.0$, $\mu_2 = 24.3$, $p > 0.05$), and mean values for Spain and the USA ($\mu_1 = 24.0$, $\mu_2 = 23.5$, $p > 0.05$) do not differ significantly, which indicates a similar attitude of tweeting on weekends. Brazil and the USA users tweet less than German users on weekends.

RQ1.1.6: Friends. The hypothesis $H_{6(a)}$ could not be supported since the multi-active users have a smaller number of friends compared to other user groups, on average. Moreover, the tests also do not support the hypothesis $H_{6(b)}$, since linear-active users (USA: $\mu_1 = 400.5$, Germany: $\mu_2 = 356.2$) mostly have greater means of the number of friends compared with other user groups. On the country-level, means for the groups of Spain and Japan ($\mu_1 = 335.4$, $\mu_2 = 337.5$, $p > 0.05$), Spain and Germany ($\mu_1 = 335.4$, $\mu_2 = 356.2$, $p > 0.05$), Japan and Germany ($\mu_1 = 337.5$, $\mu_2 = 356.2$, $p > 0.05$) do not differ significantly.

RQ1.1.7: Followers. Similarly, the hypothesis $H_{7(a)}$ and $H_{7(b)}$ cannot be supported, since the multi-active users have a smaller number of followers compared to other user groups, while linear-active users have greater number of followers compared with other user groups, on average. On the country-level, users from Spain and Japan ($\mu_1 = 296.4$, $\mu_2 = 318.3$, $p > 0.05$), users from Brazil and Japan ($\mu_1 = 335.8$, $\mu_2 = 318.3$, $p > 0.05$) do not differ significantly in the number of followers they have on average.

RQ1.1.8: User Mentions. The hypothesis $H_{8(a)}$ can be supported, since reactive users from Japan indeed mention other users the least, on average, compared to other user groups on the cultural-group and country-group levels. The hypothesis $H_{8(b)}$ can be supported, since multi-active users have greater means for user mentions compared to other users, on average. On the country-level, however, German users mention other users more than users from Brazil ($\mu_1 = 65.8$, $\mu_2 = 57.9$, $p < 0.001$).

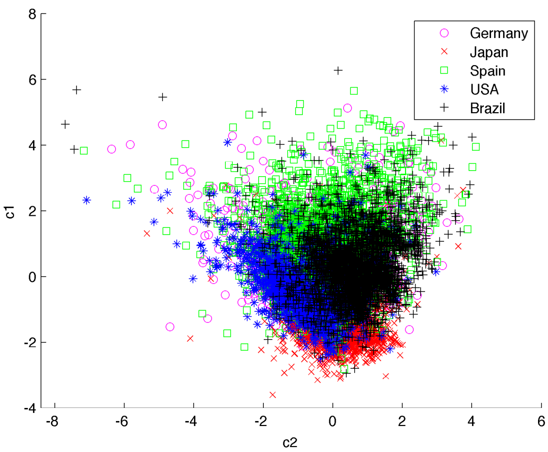
RQ1.1.9: Replies. The hypothesis $H_{9(a)}$, stating that reactive users from Japan should have more replies on average compared to other cultural user groups, can be accepted at the very high significance level. On the country-level, users from Japan ($\mu = 27.2$) behave similarly to users from Spain ($\mu = 27.5$) and the USA ($\mu = 26.2$). The hypothesis $H_{9(b)}$ can also be accepted since the average number of replies from the multi-active users is lower compared to other users. On the country-level, however, users from Spain replied more on average compared to users from the USA ($\mu_1 = 27.5$, $\mu_2 = 26.2$, $p < 0.05$). Statistics show that users from the USA reply less actively compared to other users, except Brazil ($\mu = 22.0$).

RQ1.1.10: Retweets. The hypothesis $H_{10(a)}$ can be accepted at the very high significance level. Reactive users from Japan have a smaller number of retweets on average compared with other user groups on the culture-group and country-group levels. The hypothesis $H_{10(b)}$ is also supported in our experiments, showing that multi-active users retweet the most compared to other users in our dataset. On the country-level, however, Brazilian users ($\mu = 14.3$), belonging to the multi-active culture group, retweet less (not significantly) compared to users from the USA and Germany ($\mu_1 = 15.0$, $\mu_2 = 14.9$), which are linear-active. Overall, users from Brazil, Germany, and the USA exploited the retweeting functionality in a similar way. Spanish users retweeted the most ($\mu = 23.2$) compared to other users groups.

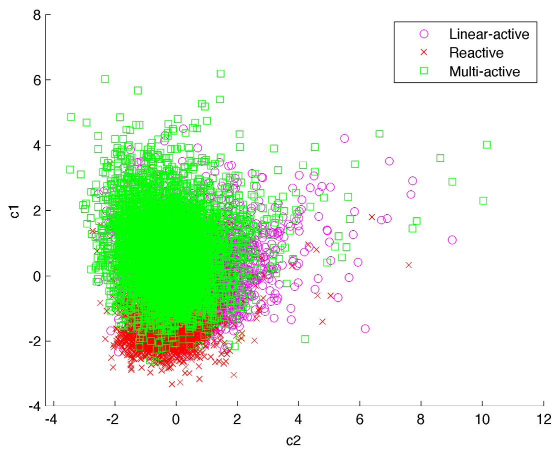
Overall, for all our tests on the culture-level shown in Table 6.2, the calculated p statistic was less than 0.001, indicating very highly significant differences between user groups on the culture-level. This corresponds to the chance of 99.9% that mean values significantly differ. Country-level tests indicate that mean values of features for country groups differ significantly in the majority of cases. Country groups of Spain and the USA, Brazil, and Japan have comparable means of hashtags and URLs usage. Spain and Brazil, Spain and the USA user groups have similar mean values of a number of tweets published during weekends. Spain and Japan user groups have comparable values of mean values for number of friends and followers. Spain and Germany, Japan and Germany have comparable means of a number of friends, while Brazil and Japan have comparable means for a number of followers. Spain and Japan, the USA and Japan have comparable means for a number of replies. Brazil and the USA, Brazil and Japan, Japan and the USA employ retweets similarly.

6.3.2 User Group Mean Distances

The Multivariate ANalysis of VAriance (MANOVA) is performed to understand whether there is a statistically significant difference between the variable means when having a set of response variables [115]. The MANOVA's necessity is discussed in comparison to multiple ANOVAs in respect to the interdependence of the analysed variables (their "redundancy"), when considering the relative importance of the variables, and when guided by the study design [115].



(a) Country Groups



(b) Cultural Groups

Figure 6.3: Country and cultural dimension clusters (MANOVA and canonical variables computed in MATLAB)

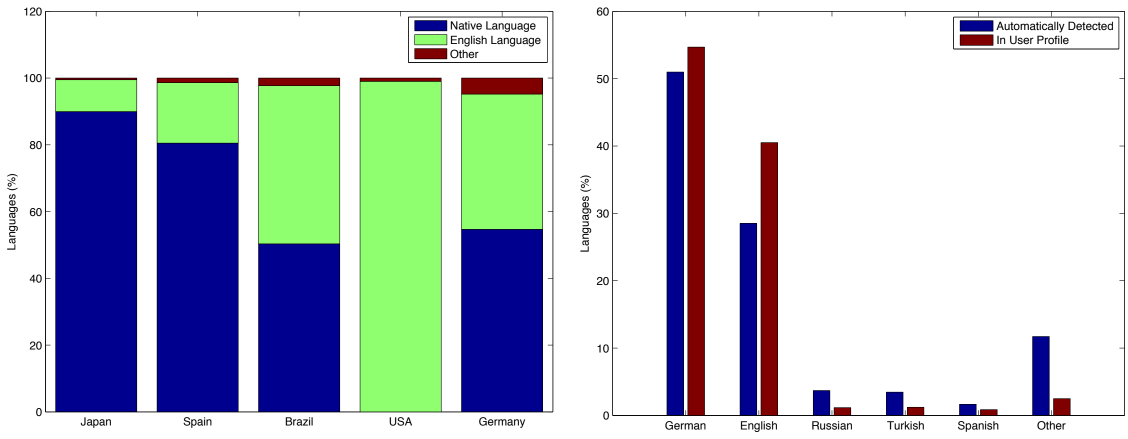
Our research questions do not pertain to the usage of MANOVA analysis. However, we are interested to overview the separation between country groups when considering the ten Twitter-based features. Based on the Multivariate Analysis of Variance, we draw scatter plots showing clusters of user groups by countries and cultural user groups in Figures 6.3 (a) and (b) respectively. The scatter plots help to visualise the differences between the user groups. Two canonical variables are used to distinguish between user groups. They are calculated from the means of the feature values analysed.

The first canonical c1 variable helps to separate clusters for the country-level user groups of Spain, Japan, the USA and Brazil. As can be seen from Figure 6.3 (a), the clusters for the user groups Spain and Japan are separated vertically, while user groups from the USA and Brazil are located on about the same level.

On the culture-level, c1 helps to separate reactive users group depicted in the red cluster below from other two clusters, multi-active users and linear-active users. This indicates that reactive users from Japan behave differently on Twitter when considering the features set analysed.

Similarly, the canonical variable c2 helps to separate user group clusters on the horizontal axis. On the country-level, c2 variable helps to distinguish clusters for users from USA and Brazil on the horizontal axis. On the culture-group level, c2 assists in separating multi-active users from the linear-active users. Figure 6.3 (b) demonstrate that the feature set enables a relatively good separation between reactive and two other cultural user groups. It is noted, however, that the multi-active and linear-active user group clusters overlap considerably.

Next, we calculate mean distances between user group means shown in Tables 6.4 and 6.5. As seen from Table 6.5 showing distances between each pair of group means for the mix of the aforementioned features, the distance between linear-active groups and multi-active group means (1.09) is much smaller than the distance between multi-active and reactive groups (4.06). For instance, the distance between German and Spain means is about 0.9, while the distance between the Spain and Japan is about 4.65 as seen from Table 6.4.



(a) Native and English languages in user profiles (%)

(b) Foreign languages in German profiles and content (%)

Figure 6.4: Fraction of languages automatically detected in the tweets and Twitter user profiles (considering the English language popularity in Twitter profiles discussed in [237], we define here the most used language in the country of tweet origins as the Native language, for instance, for tweets originating from Germany, we define the German language as Native.)

Country-level				
Test	Features	Resubstitution Error	Nodes	Cross-validation Error
1	$LANG$	0.22	6	0.22
3	DEF	0.17	51	0.42
5	$DEF + LANG$	0.02	680	0.06
Culture-level				
Test	Features	Resubstitution Error	Nodes	Cross-validation Error
2	$LANG$	0.17	4	0.17
4	DEF	0.10	47	0.29
6	$DEF + LANG$	0.01	511	0.04

Table 6.3: Resubstitution (applying trained classifier for each data point in train set) and cross-validation (unbiased estimate, applying trained classifier on separate dataset 10 times) error rates for predicting user groups with decision tree classification. Feature sets include the DEF - features analysed in the section 6.3, $LANG$ - language in the user profile, $DEF + LANG$ - combination of previous two (feature frequency distributions and importance shown in Appendix C.3.1, confusion matrices of these tests using Decision Trees and two other classifiers created using Microsoft AZURE platform are shown in Appendices C.2 and C.3.2 respectively).

	Japan	Spain	USA	Brazil
Germany	**3.51**	0.90	1.7	1.20
Japan		**4.65**	2.19	**3.17**
Spain			**2.23**	1.12
USA				2.74

Table 6.4: Distances between country group means

	Reactive	Multi-active
Linear-active	**2.54**	1.09
Reactive		**4.06**

Table 6.5: Distances between cultural group means

Interestingly, distances between both linear-active groups (the distance of 1.20 between Germany and Brazil, and 2.74 between the USA and Brazil) and Brazil are larger than between the linear-active groups and Spain (the distance of 0.9 between Germany and Spain, and 2.23 between the USA and Spain). This coincides with the Lewis model in that Spain is more close to the linear-active triangle corner than Brazil, considered the "extreme" multi-active country. Therefore, we can conclude, that the linear-active and multi-active user groups are more similar in their behaviour, while reactive users behave differently on Twitter in respect to the analysed features.

6.3.3 Prediction Quality

To assess the quality of user profiling based on the analysed feature set, we created six decision tree classification models [11]. The first two models (1 and 2) were created based on the language defined in the user profile. However, languages specified in the Twitter user profile could be misleading. For instance, in our dataset, a large fraction (about 40%) of users from Germany specified their preferred language as "English", as shown in Figure 6.4. This is why we also created classification models (3 and 4) based on the selected features set while excluding languages defined in the user profile. Models 5 and 6 were created by combining features set and languages specified in the user profile.

The classification models enabled us to predict users belonging to a user group on country-level or culture-level. The classification models were assessed by calculating

[11] The decision tree selection is justified in Appendix C.1

resubstitution error rate and testing error rate. For cross-validation, we split our sample into ten almost equally sized parts used for finding out the testing error rate. Table 6.3 shows resubstitution errors, number of terminal nodes for pruned trees, and cross-validation errors for aforementioned tests and feature sets defined. As it can be seen from the table, when the profile information on languages is not available, the DEF features set can be used to predict a user belonging to cultural dimensions or one of the five countries, analysed with a relatively high cross-validation error rate of 0.29 and 0.42, respectively. This indicates that the DEF features set might be further extended with languages, other features when available in the profile or tweets content of the user. The combination of the DEF and $LANG$ feature sets enables the lowest cross-validation error for culture-level and country-level classifications. The cross-validation error for the feature set $DEF + LANG$ decreased by 73% and 77%, on country and cultural-levels respectively, compared to the cross-validation error when employing only languages defined in the user profile. We, therefore, concluded that we could exploit the analysed Twitter features for predicting user origins on a country and geographic region levels (RQ1.2).

6.3.4 Interpretation of Results and Discussion

Cultural differences and Country-level Similarities. Based on descriptive statistics and comparison of mean values of features for different cultural groups, we found distinct differences between the reactive user group and other user groups. Japanese belong to the reactive users group, and they share the least of hashtags and user mentions. Japanese reply, however, more than other user groups, with the exception of Germans. Japanese retweet less compared with other user groups. This can be explained by their good listening skills and “high-context” orientation as explained in [148]. Japanese users also tweet the least from different geographic locations. Moreover, we detected the least of foreign languages in the content published by reactive users compared to others. Japanese also tweet more on weekends compared with other user groups.

Interestingly, even though we initially hypothesised that multi-active people as more people-oriented persons might have larger social networks of friends and followers, tests showed that linear-active users from the USA and Germany have, on average, more followers [119]. They also have more friends compared to other user groups, except for users from Japan, for which they show a comparable mean value. Linear-active users also generally share more URLs compared with other user groups. Interestingly, German users belonging to the linear-active group have the greatest mean for replies compared to other users. Overall, linear-active users share also more hashtags compared with other groups but Spain. The means of hashtag usage are similar for Spanish users and users from the USA.

Moreover, multi-active users have similarities with linear-active user groups and are therefore difficult to separate. Considering the multi-active users group, Spanish refer the most to other users (mentions usage) and are quite similar in their behaviour with the USA group, while Brazilians share fewer links, and only refer more to other users than Japanese. For multi-active users, we detected more foreign languages on average compared with linear-active users. They also have a smaller number of followers and friends compared to others.

Our findings agree with the study of [78] indicating that persons from Eastern countries are less individualistic, refraining from the usage of hashtags. In [193], users from South Korea and Japan have a smaller fraction of hashtags in their tweets. Our experiments also correspond with findings of [193] in that Japanese persons employ fewer user mentions than persons from Western countries. Our findings reveal that Japanese users retweet the least, which corresponds with [193], while they reply the most. This corresponds with [148] stating that reactive persons are generally good listeners and prefer in-depth content.

Our findings also correspond with the study of Lewis [148] in that linear-active Western persons are “data-oriented”. We found a similar pattern of URLs usage as in [193] where users from the USA share the most URLs compared to others. In opposite, as explained in [148], multi-active and reactive persons are “people-oriented”. Our experiments support this idea, since persons from Spain, Japan and Brazil share less URLs compared to the “data-oriented” persons from the USA and Germany.

Multi-active persons are described as loquacious in [148], in our experiments, Brazil and Spanish users also employ the most foreign languages. To summarise, some of the findings correspond with the previous studies by [193] and [148]. This indicates that we found similar microblogging culture-specific behaviour patterns even though working with different data-sets of Twitter users. It, therefore, appears that human communication in social networks could be influenced by cultural differences, which could be further explored in future studies to facilitate better user experience in social or virtual environments.

Cluster Analysis. The cluster analysis showed that the distance between Spain and Germany is smaller than the distance between clusters of Germany and the USA. Also, linear-active users behave similarly to multi-active users when analysing clusters formed from the multivariate analysis of their variances for the analysed features. The user group from Germany is difficult to separate from the user groups of the USA, Spain and Brazil. We explain it by possible cultural similarities between these user groups and how they behave on Twitter. It is also reasonable to assume that this could be explained by the peculiarity of our dataset or in relation with the Lewis model. This is why we cannot confirm the strict relation with the Lewis' model.

Moreover, Lewis [148] stated that Spanish people coming from different regions might behave very similarly to linear-active people in the sense of productivity. The geographic proximity also has a substantial impact on personality across cultures [15]. This implies that there are more variables and relationships which might be considered for creating cultural user models based on microblogs.

For instance, the features set can be further extended with topics derived from the tweets content and user opinions mined in a process similar to works such as [172] and [272]. Cross-cultural topics analysis in tweets can be considered as a direction for future research.

In addition, the study by Gao with co-authors [78] informs us about different fractions of positive posts for users of Sina Weibo and Twitter. This is why more features, like for instance emoticons could be added to the classification model to reflect differences in expressing feelings and moods. Real-life communication differences between people of different cultures as explained by sociological models thus can be further analysed in the context of microblogging behaviour and self-expression. A possible research direction could be to investigate how we could mine affective states from microblogs and how they reflect real-life communication patterns.

Localisation and Adaptation Assumptions. Nevertheless, microblogging patterns on the country-level still can reveal users' attitudes on how they use the Twitter functionality. The insight that linear-active users from Germany and the USA tend to share more URLs and hashtags, have a larger contact network might suggest that the related microblogging functionality can be further enhanced for these users. For instance, a reply button functionality could be more visible for reactive users willing to participate in a more substantial dialogue, instead of providing a button for retweeting, which might be preferred by users from Brazil, Spain and the USA.

Furthermore, the distance between clusters for linear-active and multi-active users is about 1, while the distances between reactive and multi-active, between reactive and linear-active user clusters is about 4 and 2.5 respectively. It seems that the features analysis shows us that reactive users stay apart from the other two groups. As it was suggested by Lewis in [148], marketing efforts should not neglect reactive and multi-active persons, which worldwide are more than linear-active persons. The design and functionality of social networking websites and other applications can be tailored to the particular cultural user groups to reflect their preferences. In this sense, our findings agree with [254] and [257] on localisation benefits for social network services targeting users from different cultural origins.

Data Collection and Experimental Setup. It is important to note that our study was based on the users sharing their geographic locations. We have restricted our crawling process to the big cities in five selected countries. As it is advised by [257], more in-depth research is needed to analyse more countries and social networking services. We agree with this and in future work, we aim to extend our framework with more countries/users to allow analysis on a larger scale.

Our original dataset included on average more than 600 tweets per user. For building individual user profiles, we considered however only 100 tweets, since otherwise, we would only be able to model less than 300 users from the USA user group. Therefore, following our assumption that classification performance increases given more users, we selected 100 tweets as a starting point in our experiments based on more than thousand users per country group. In further experiments we plan to extend our users dataset and investigate the number of users/tweets required to build representative user profiles for modeling cultural origins. This would allow to better understand how classification performance scales with number of users and tweets included in the user profile. We believe that increasing the number of users would enable better prediction outcomes in the classification experiments we performed. Based on the previous findings in [160, 161] indicating the consistency of psychological behaviour traits in different gender and age groups within national borders, we do not distinguish between different social groups. However, further research might confirm the reliability of this assumption while analysing user behaviour on Twitter or other microblogging services across the cultures.

Moreover, our experiments on classifying user profiles showed that we could employ classification methods such as decision trees to classify users into particular user groups on the culture and country levels. The analysed feature set extended with the language defined in the Twitter profile of the user enables a low cross-validation error rate. However, when language information is not available, the language can be inferred automatically from user content. Alternatively, more features derived from the user profile/content can be further analysed to improve quality of users classification, which can be performed using other methods such as logistic regression or ensemble classifiers. In further work, we aim to implement and analyse other classifiers in order to facilitate separation of users from linear-active and multi-active countries. Such classification can be further exploited by adaptive applications when knowledge on user cultural background is needed.

Nevertheless, whilst it is challenging to assess adaptation outcomes next to a statistical observation [197], user modeling efforts could be beneficial for improving user experience. Previous studies have shown that users of adaptive applications can benefit from adaptive functionality features. An empirical study by Strachan with co-authors [234] has shown that simple user modeling introduced into a commercial application influenced positively user perception of software capabilities. Forbes-Riley and Litman [75] found a positive correlation between learning outcomes and adaptability to a learner state of uncertainty in a dialog-based tutoring system providing adapted feedback to learner answers.

6.4 Conclusions

In the foregoing, we analysed microblogging behaviour on Twitter for user groups from Germany, USA, Spain, Brazil and Japan. We found that Japanese users behave very differently from the rest of the user groups. In comparison, they tweet more on weekends, reply more and share the least of hashtags and user mentions. In contrast, users from the USA and Germany generally share more URLs and have more friends compared with the other user groups. Users from Spain and Brazil stay apart in a way that they have some similarities with the rest of groups but are difficult to differentiate when using the analysed set of features. Multi-active users, however, appear to employ more foreign languages than others.

We reflected on the results with the help of the sociological model by Lewis. Whilst it was not possible to explicitly map cultural-related communication patterns to microblogging behaviour on Twitter, some of the derived microblogging patterns enabled us to distinguish between different cultural groups on Twitter. Based on the found microblogging patterns, we proposed an approach of culture-oriented user modeling that considers cultural/country differences of the users. The information on user microblogging activities, preferences for information sharing and/or dialogues can be further exploited for designing adaptable applications which suit to user needs based on her cultural/country origins. In the following chapter, we employ a similar approach for inferring user countries for further analysis of user communication flows in Twitter. Since we aim at a larger set of countries, we might need to reconsider the feature set. For instance, we are interested to find out whether user follower network might be used for finding out the user locality. The understanding of personal communication preferences might be useful for creating networking friend recommendations, or content recommendations tailored to user locality-specific preferences.

Chapter 7

Communication Preferences

"Culture is communication and communication is culture. "

- Edward Twitchell Hall [22]

This chapter is based on the publication "Cultural and Geolocation Aspects of Communication in Twitter", co-authored with N.K.Taylor and Y.Jing, and presented at the ASE Social Informatics conference at Harvard in December 2014.

Web applications exploit user information from social networks and online user activities to facilitate interaction and create an enhanced user experience. Due to privacy issues, however, it might be challenging to extract user data from a social network, in particular, location data. For instance, information on user location depends on users' agreement to share own geographic data. Instead of directly collecting personal user information, we aim to infer user preferences by detecting behaviour patterns from publicly available microblogging content and users' followers' network. With statistical and machine-learning methods, we employ Twitter-specific features to predict country origin of users on Twitter with an accuracy of more than 90% for users from the most active countries. We further investigate users' interpersonal communication with their followers. Our findings reveal that belonging to a particular cultural group is playing an important role in increasing users responses to their friends. The knowledge on user cultural origins thus could provide a differentiated state-of-the-art user experience in microblogs, for instance, in a friend recommendation scenario.

7.1 Introduction

Social networking sites such as Twitter microblogs allow users to communicate with their online friends and share information in real time. Some web and mobile applications require information on user location and origin to provide users with location-specific information, like recommending places of interest (Foursquare) and restaurants (Opentable) in close proximity. A user profile containing user geographic locations, on which recommendations can be provided may also impose possible privacy threats. This is why a majority of Twitter users avoid sharing their accurate locations [104].

However, despite user efforts to hide or obscure their whereabouts, there are methods to identify user origins based on content [100, 85, 40], social profile metadata [100, 103], and from other social networking sites, in which users' data is gathered [129]. Besides location-related cues, microblogging activity patterns can reveal users from different origins, which could be used for indirectly inferring origins as we discussed previously in chapter 6.

We do not aim at finding user locations on a city or state level, which is addressed by Mahmud with co-authors [156]. Striving to preserve user privacy, we abstract from mining accurate location-specific information. We limit ourselves to a country-level or a "cultural group" comprising a group of countries. User similarities and differences found in the profile metadata, contact networks, content and microblogging behaviour can be employed as a proxy for finding user origins, whether they are country or culture-related. In this chapter, we discuss our contributions to social networking research as follows:

- Investigation into the predictive value of Twitter user-related and social-network related features in experiments for predicting user origins.
- Analysis of user communication preferences in Twitter on the example of followers' responses to their friends.

In the following, we outline related work in the scope of Twitter location detection, cultural aspects playing a role in adaptation and personalisation. Next, we describe our research methodology and experimental setup. Then, we summarise the results in user origin and response prediction experiments. We also discuss benefits and limitations of the proposed approach and possible areas of improvement. We conclude with research findings, how our findings could be used in the development of adaptive social web applications.

7.2 Background

There are several research domains investigating adaptation and personalisation outcomes considering cultural backgrounds of users. In e-learning, Olaniran [180] stated that the cultural context of a learner might impose requirements on technology usage, selection of media and the style of interaction between students and instructors. The learning material presentation methods impact the performance of students from different cultural groups [262]. To improve the experience of e-learners, it is paramount to distinguish between different cultural preferences [180].

In a recommendation system context, Silva with co-authors [225] analyses recommendation prospects for urban planning in accord with user cultural similarities on Foursquare and user preferences. Cultural behaviour differences in user activities on Twitter were investigated in [80], suggesting to exploit such differences for building responsive communication applications. Other possible applications include friendship recommendations based on suggestions by the network [83] and a community detection approach based on user interactions on Twitter [87]. Microblogging response prediction with focus on English-speaking users was studied in [17], which analysed the importance of specific terms occurred in tweets, previous response ratios and also number of user and follower links in the social network. The importance of a social relationship between Twitter users on their responses prediction was revealed in [210].

Overall, the above mentioned research points out cultural behaviour differences between users online. It remains however unclear whether knowledge on the user cultural background would help in improving recommendation performance, and whether such recommendations would outperform country-specific recommenda-

RQ2.1	Could we exploit user contact network, i.e., friends, for predicting user cultural origins?
H2.1	The information on country location derived from user tweets' metadata and respective metadata of the followers is insufficient for providing cues on user origins for the majority of users ($\geq 75\%$).
RQ2.2	What microblogging features (user-related and friend-related) are the most prominent in revealing user cultural traits in Twitter?
H2.2.1	A user's contact network can assist in improving country prediction.
H2.2.2	User microblogging patterns can assist in further improving locality prediction.
RQ2.3	Could we find communication preferences in respect to user cultural origin?
H2.3.1	User and follower's country locations' and language match are amongst the most important prediction parameters for user responses.
H2.3.2	User's influence is significant in predicting her follower responses.

Table 7.1: Research questions and hypothesis
(see all project-related objectives and questions on page 69)

tions. A further in-depth investigation into follower and friend relationships on Twitter could shed light on location-specific cultural aspects playing a role in online communication. For predicting user responses, we further evaluate the inclusion of geographic and culture-specific aspects of the users involved. For this, we examined the user and follower-related features enabling to identify user country and culture-level origins, which we further exploited in a user response (reply or retweet cases) prediction experiment.

7.3 Methodology

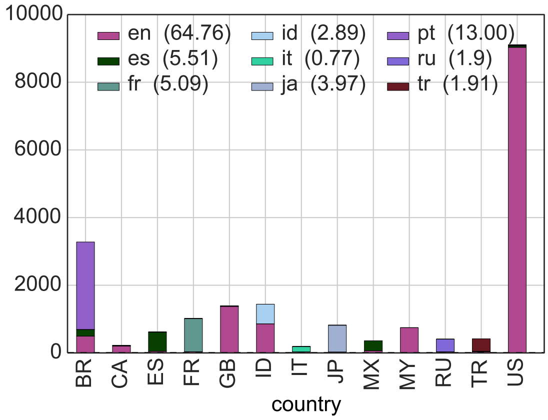
The main aim of this chapter is to analyse user communication in Twitter and to get insights into friend-follower relationships to predict follower replies. It seems reasonable to assume that user geographic locations, cultural origin, and language might play a role in the follower interests reflected in reply or retweet messages.

We also analysed the relative importance of other Twitter-specific features. For instance, a number of user followers and friends, which ratio is often explained as an "Influence" on other users might help us to detect interesting users to follow. Also, we assess user-related and follower-related features for predicting a user whereabouts and the interest of the followers. We present our research questions and hypothesis

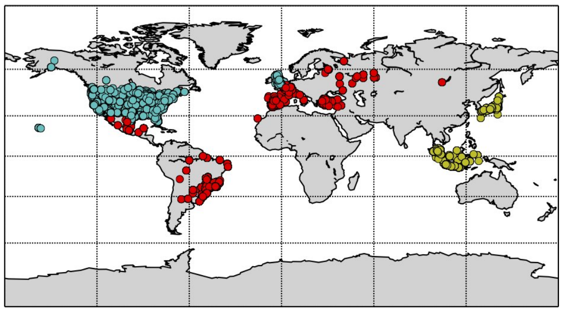
Country (Code)	Dimension	Country (Code)	Dimension
Brazil (BR)	MA	Mexico (MX)	MA
France (FR)	MA	Russian Federation (RU)	MA
Spain (ES)	MA	Italy (IT)	MA
Turkey (TR)	MA	Indonesia (ID)	RE
The United States (US)	LA	Japan (JP)	RE
Great Britain (GB)	LA	Malasya (MY)	RE
Canada (CA)	LA		

Figure 7.1: Countries and assigned dimension codes

in Table 7.1. To create three main cultural groups, including MA, LA and RE, we combine users from several countries. With human annotation, dimension codes where assigned to the countries which were analysed, as seen in Figure 7.1.



(a) Languages (overall percentages are in brackets above) and countries (ISO codes), the vertical axis shows the number of users



(b) User locations and and associated cultural dimensions in Lewis Model [148] (blue: linear-active, yellow: reactive, red: multi-active)

Figure 7.2: Top languages and geographic locations

7.3.1 Data Collection

Selected Users. Using the Twitter Streaming API, we collected a sample of of 4250109 tweets [1] published by 3198307 users in the period from 17th to 18th of March 2014 by listening to the Public Twitter Sample Stream. Only about 2% out of these tweets were provided with geographic locations. From this tweets sample, we randomly selected 20,000 users with published tweets containing geographic locations as seen in Figure 7.2. The geographic information available in the tweets'

[1] Appendix F summarises the top ten countries with geotagged tweets from this sample

metadata helped to reduce pre-processing time. To accurately determine a user's origin, it was paramount to decrease as much of possible the number of users travelling or residing in countries other than their country of origin. For this, we introduced a parameter α, which equals one when the user language defined in the user profile matches with the top first native language related to the user's country and zero otherwise. This requires a large number of users when training our classification models for improving accuracy when identifying users and their respective origin.

User Profiling Features. For the selected users, we followed their tweets, replies and retweets of their followers in the period from 18th to 25th of March 2014, in total amounting to 2,853,719 tweets collected, out of which around 31% were geographically tagged and provided with information on countries of origin. Using these geo-tagged tweets, we created feature vectors representing summaries of the selected user microblogging activities; these aggregated feature vectors were stored in the table "Profiles" (summary in Table D.3). The "Profiles" table consists of 13289 pre-selected users (from the initially selected set of 20,000 users who tweeted during the data collection time) defined as tweeting [2] from the top most active countries in our sample dataset, namely the USA (US), Brazil (BR), Indonesia (ID), Great Britain (GB), Turkey (TR), Japan (JP), France (FR), Spain (ES), Malaysia (MY), Mexico (MX), Russia (RU), Canada (CA), Italy (IT) and where content from followers was available (they got replies or retweets during the time of data collection). We exploited the following features for detecting user country origin:

- LANGUAGE: user language from user's Twitter profile.
- BEHAVIOUR: features set included features related to a single user's activities on Twitter:

    1. **Tagging**: number of Hashtags divided by the sum of Hashtags and Uniform resource locators (URLs) occurring in user-generated content (in our pilot test we found out that the combining Tagging and URLs sharing activities into one measure leads to better results in our experimental

[2] sending Twitter messages

setup); denotes the user's preferences towards tagging and sharing content activities.

2. **Languages**: number of different languages detected from the user content[3], normalised by division of the mean values of languages employed by all users.

3. **Weekends**: number of tweets published on weekends divided by the number of tweets published by the user; denotes frequency of posting during weekends.

4. **Replying**: number of user replies divided by the total number of user replies and retweets.

5. **Mentions**: defines user preferences for sharing information on other users as compared to sharing Hashtags and URLs; calculated from the total number of user mentions divided by the sum of URLs and Hashtags. This feature reflects users' focus on people or organisational activities as described in Lewis [148].

6. **Mobility**: denotes the number of different country occurrences in the user tweets' metadata divided by the mean value of the number of country occurrences in the tweets of all users.

7. **Timezones**: number of different time zones in the tweets' metadata of a user.

8. **Influence**: number of Followers divided by the total number of Followers and Friends.

- *PLACE.* The text string created by joining strings of the language code defined in the user profile, most used Time zone and Location found in the tweets' metadata.

- CONTENT: for each user we consider the textual content of only one tweet.

[3]Python library "langid" at https://github.com/saffsd/langid.py is fast and robust for Twitter messages while comparable with or outperforming in its accuracy of language detection of other solutions including LangDetect (LangDetect we used in the previous chapter, page 91, while using Java programming language) and TextCat [154].

- LOCATION: location field found in the user metadata.
- FOLLOWERS features set consists of features extracted from the user follower networks.

  1. **FCountry**: the country mentioned most in followers' metadata.
  2. **FCountries**: the number of different countries referred to in the followers' metadata.
  3. **FLanguage**: the language most mentioned in the followers' tweets' metadata.
  4. **FLanguages**: number of different languages found in the followers' tweets' metadata.
  5. **FTimezone**: time zone most referred to in the followers' metadata.
  6. **FTimezones**: number of different time zones referred to in the followers' metadata.
  7. **FInfluence**: number of Followers divided by the total number of Followers and Friends for each of the user Followers, further taking the mean value for the user we follow.

For determining a user country location, we employed Twitter metadata and Twitter specific elements. We did not apply named-entity recognition algorithms, since they are resource-demanding and some named entities may be quite ambiguous in distinguishing from other words [177, 171]. The initial feature choice was guided by the literature sources and our experimental system design. In addition to content and user-related features, we also included follower-related features, to examine their importance in relation to country detection performance. In further experiments, we also examined the user responses in friend-follower relationships.

7.3.2 Country Detecting Classifier

Each of our initially selected users had an associated country code (string with ISO 3166-1 alpha-2 country code) found in the user metadata of tweets. We used these users for training our classification model based on the features described in this section.

To each of LANGUAGE codes, we assigned a numerical value, while the BEHAVIOUR feature set included only numerical values. In the FOLLOWERS feature set, we coded FCountry, FLanguage, FTimezone as numerical values, while the rest were computed as integers (FCountries, FLanguages, FTimezones) or real values (Influence). When dealing with numerical values, for building our classification models, we exploited the Decision Trees Classifier, which also allowed to consider the importance of a feature.

When dealing with language features extracted from the user tweets (CONTENT, LOCATION, and $PLACE$), we created pipelines performing the following steps:

- Convert text data such as tweets' content or location field to a matrix of token (strings of at least two alphanumeric characters) counts;
- Convert the count matrix into its normalised Term Frequency - Inverse Document Frequency (TF-IDF) statistics;
- Employ the Multinomial Naïve Bayes classifier for predicting user countries based on the normalised TF-IDF sparse matrix.

For evaluating our country classification results, we performed a three-fold Cross-Validation (CV), and employed Accuracy and F1-measure [67]. For creating our training and test samples, we split the "Profiles" data table into fractions of 75% and 25%.

7.3.3 Communication Patterns and User Responses

Based on the "Follow" data table, we created the "Communication" data table (see Table D.4) representing 107,960 user tweets, replies, and re-tweets, from which about 38% were geographically tagged and published by the pre-selected users and their followers. We then analysed user interactions among different user groups. Next, we created and evaluated a response predicting classifier based on the ratings computed using the “Communication” data table as follows:

- For each of the pre-selected users, we computed the number of their followers’ retweets and replies, which we exploited for defining an “Interestingness” value of the initially selected users for their followers. The “Interestingness” equals to 0 when no replies and retweets for a particular user, and to 1 when there is a maximal number of retweets and replies for the user and follower communications. We thus labeled each pre-selected user to define his or her “Interestingness” for a particular follower.

- We predicted user interest towards their friends by training our decision tree and logistic regression models based on the aforementioned features and a set of features denoting users’ matches in the language usage and location.

We also added binary variables such as “LangMatch”, “CountryMatch”, “DimMatch” (True when user and followers’ profile languages, Countries, and Dimensions match), “FLangMatch” (True when user followers’ and followers followers’ profile languages match), “FCMatch” (True when user followers’ and followers followers’ Country match), ”FTimezMatch” (True when user followers’ and followers followers’ time zones match).

When a followers’ location was unknown, we employed the previously constructed classification models “PLACE” and “LANGUAGE”. The parameters “CountryMatch” and “DimMatch” were set to True when friends’ parameter matched with the parameter of the follower based on one of the three values: value from the profile information, parameter’s value derived using “PLACE” or “LANGUAGE” country detection models, or employing “LANGUAGE” model for detecting a dimension value directly. For instance, when the inferred cultural dimension of the follower was “Linear-Active” and the cultural dimension of the followee was also “Linear-Active”, “DimMatch” was set to True.

In brief, we used Tagging, Languages, Weekends, Replying, Mentions, Mobility, Timezones, Influence, FCountries, FLanguages, FTimezones, FInfluence (features described on pages 120 - 122), and binary features including CountryMatch, DimMatch, FCMatch, FLangMatch, FTimezMatch, LangMatch for predicting follower interest towards their friends (the "Interestingness" value).

7.4 Results

7.4.1 Explicit Locations in Metadata

We were interested in evaluating how useful country-related information extracted from user tweets' and followers' tweets is. The aim was to find out the required amount of Twitter-content to infer country-locations of users in our dataset and assess the possibility of detecting user origins based on the metadata of tweets of the user and followers.

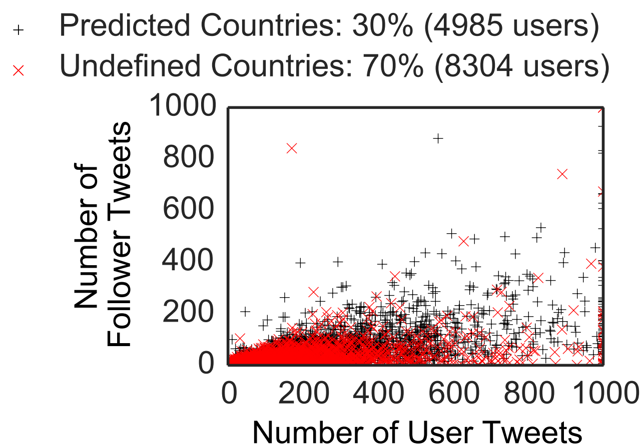


Figure 7.3: Countries detection test using followers' locations in their metadata (on axes x and y we show the number of tweets published by initially selected users and their followers respectively during our data collection; each point on the figure depicts a fellowship link between these twitter users; for each pair of the initially selected user and his/her follower, we found the most frequent country location based on her/his tweets' meta-data; when these top country locations were found and matched for the selected users and their followers, we marked the respective points as "+" signs, otherwise, we marked the points as "x" signs)

Our first hypothesis states that metadata of user and follower's tweets are not sufficient for explicitly inferring country locations for a majority of selected users. The average number of tweets per user is about 119 tweets (standard deviation 156 tweets), while the number of friends and followers is about 590 (st.dev. 2731 friends) and 894 (st.dev. 7598 followers) respectively per user on average. Even though the fraction of geo-tagged tweets from our "Follow" data table exceeds the randomly selected tweets from the "Sample" data table in more than 15 times, reaching around 31%, the amount of tweets collected during one week was insufficient to infer country-specific information for about 70% of the users (see Figure 7.3).

In spite that geographically-tagged tweets occur only in about 2% of cases in our "Sample" data table, the location field enabling users to specify their locations arbitrary in text was filled in 54.6% of cases. However, here we might disregard the location field (further examined below in Table 7.2) and usage of geocoding services due to the ambiguity and often use of the location field by Twitter users for personal or humoristic comments, as stated in [104] in more than 30% of cases. Therefore, we assert in our H1.1 hypothesis, that for the majority of cases the country location metadata is insufficient for inferring user countries based on followers' and user tweets metadata alone.

7.4.2 User Country Prediction

Next, we analysed how the different combination of features can be useful for predicting user countries (or countries from which users send their messages on Twitter).

User Data. We aimed to achieve an accuracy of above 46% of country prediction when considering the majority class classifier's threshold, since about 46% of users in the "Profiles" data table originated from the USA, with English defined in their user profiles. Moreover, in our dataset some of the countries such as Indonesia (ID) and Malaysia (MY) have a large fraction of English language users, even though English is not the native language. This is why some of the users are misclassified if only the language defined in their user profiles is considered. However, our findings revealed that the LANGUAGE classification strategy outperformed all other User-related strategies, except $PLACE$, in respect to CV accuracy, as seen in Table 7.2.

	CV Acc.		Accuracy		Precision		Recall		F1	
Feature Set	α=1	$\forall\alpha$	α=1	$\forall\alpha$	α=1	$\forall\alpha$	α=1	$\forall\alpha$	α=1	$\forall\alpha$
User-related Data										
LANGUAGE	0.88	0.76	0.88	0.76	0.78	0.70	0.88	0.76	0.82	0.71
LOCATION	0.62	0.58	0.64	0.59	0.63	0.66	0.64	0.59	0.53	0.51
PLACE	**0.91**	**0.85**	**0.91**	**0.85**	**0.90**	**0.86**	**0.91**	**0.85**	**0.90**	0.83
BEHAVIOUR+LANGUAGE	0.80	0.66	0.81	0.66	0.81	0.67	0.81	0.66	0.81	0.67
CONTENT	0.63	0.58	0.65	0.58	0.55	0.45	0.65	0.58	0.54	0.47
BEHAVIOUR	0.44	0.38	0.46	0.38	0.48	0.39	0.46	0.38	0.47	0.39
Follower-related Data										
FOLLOWERS	0.88	0.84	0.88	0.84	0.88	0.84	0.88	0.84	0.88	**0.84**
Mixed Data										
LANGUAGE+FOLLOWERS	**0.94**	**0.87**	**0.94**	**0.87**	**0.94**	**0.87**	**0.94**	**0.87**	**0.94**	**0.87**
BEHAVIOUR+FOLLOWERS	0.87	0.82	0.88	0.83	0.88	0.83	0.88	0.83	0.88	0.83
BEHAVIOUR+FOLLOWERS+LANGUAGE	**0.92**	**0.87**	**0.91**	**0.87**	**0.91**	**0.87**	**0.91**	**0.87**	**0.91**	**0.87**

Table 7.2: Performance of the country-detecting classifier

(in test/train split: 4650 training and 1550 testing instances; in cross-validation test: the original sample of 6200 instances is randomly partitioned into three equal size parts; cases with $\alpha = 1$ when languages defined in the user profile match with the mostly used language in the country of the tweet origin; across all columns we marked in "**bold**" font the top three highest values for all metrics and α levels, which revealed the most successful feature sets including LANGUAGE+FOLLOWERS, BEHAVIOUR+FOLLOWERS+LANGUAGE, and *PLACE* features set requiring overall smaller pre-processing and data collection efforts since we do not use any follower-related information)

The $PLACE$ feature is represented by a text string comprising language defined in user profile, majority time zone and location. With the $PLACE$ feature, we achieved 91% CV accuracy (three-fold CV showed the best CV accuracy performance in all our tests), which outperformed the performance of the LANGUAGE strategy. Since the location-specific information is not always accurate [104], we further analysed other feature mixes. The CONTENT strategy slightly outperformed LOCATION-based strategy in $\alpha = 1$ cases for Accuracy, Recall and F1 measures while enabling to achieve 63% CV accuracy when based on only one tweet's content per user.

The BEHAVIOUR strategy performed poorly compared with other classification strategies, however, we noticed a considerable improvement of performance metrics (improvement of Cross-Validation accuracy by 13%, Test Accuracy by 17%, Precision by 19% and Recall by 17%) in $\alpha = 1$ cases as compared with $\alpha = 0$ cases. When joining BEHAVIOUR and LANGUAGE feature sets, in all $\alpha = 1$ cases we could not exceed the performance achieved with the LANGUAGE feature alone for all performance metrics except Precision, which improved slightly by 3%. It seems that the BEHAVIOUR feature does not allow us to improve user classification into country groups when using only user-related data. Considering a relatively small number of users in our "Profiles" data table, we were not surprised to achieve only 44% of CV accuracy and 46% test accuracy when using the BEHAVIOUR feature set, which might require more data and refined features to yield similar results to the strategies analysed above.

Followers Data. In cases when user metadata was not available or deemed to be inaccurate, the FOLLOWERS strategy could compete with some User-related feature sets. We achieved more than 5% improvement in Precision and F1-measure compared to all User-related feature combinations, except PLACE. Interestingly, in a majority of test cases, we observed better performance when considering $\alpha = 1$ cases. We might reasonably assume, that the combination of user language and user country of origin is important for selecting training instances. Overall, the FOLLOWERS feature set provided a viable alternative for detecting user countries in our experiments when accurate PLACE and LANGUAGE data was absent.

User and Followers. When combining user LANGUAGE with information extracted from the followers' network, we achieved the best performance in all measures, in all our experiments for $\alpha = 1$ cases. The cross-validation accuracy of LANGUAGE+FOLLOWERS combination was around 94%. Therefore, we might accept our hypothesis H2.2.1, that the user contacts' network improves user country predictions. Despite our expectations, combining BEHAVIOUR with FOLLOWERS features did not improve classification performance compared to using only FOLLOWERS. We could not accept the H2.2.2, that the BEHAVIOUR patterns could help in improving user country predictions in our experimental settings. Overall, when using Follower-related features and LANGUAGE, we were able to outperform the accuracy when using only LANGUAGE/PLACE features and also the "Calgari" algorithm [104] using Naïve Bayes classification model with the first 10,000 terms having the highest scores based on the conditional probability of term in geographic locations such as countries or states.

BEHAVIOUR+LANGUAGE+FOLLOWERS					
Feature	Importance (%)	Feature	Importance (%)	Feature	Importance (%)
Language	100	FCountry	16.16	FTimezone	15.36
FInfluence	2.34	Weekends	2.34	Influence	2.13
Mentions	1.55	Tagging	0.98	Replying	0.73
FTimezones	0.51	Timezones	0.36	Languages	0.33
FLanguage	0.32	FLanguages	0.26	FCountries	0.15
Mobility	0.02				

Table 7.3: Relative features importance in user country prediction

Combining all non-content based features in the "BEHAVIOUR+LANGUAGE+FOLLOWERS" classification strategy enabled us to assess the relative importance of features for country detection using Decision Tree classifier. Our experimental results showed that user language, followers' majority country and time zone, followers' average influence and posting day were the most important features for detecting user country of origin, while the user mobility was the least important feature to consider, as seen in Table 7.3.

7.4.3 Predicting Follower Responses

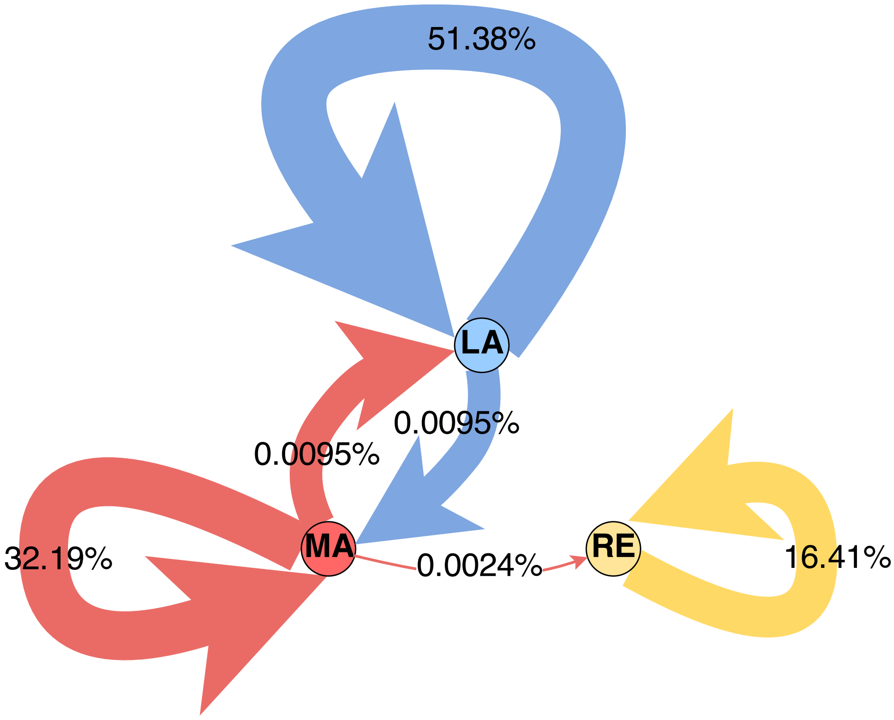


Figure 7.4: Communication between Cultural Dimensions for geographically-tagged users; LA: Linear-Active, MA: Multi-active, RE: ReActive user groups

Our analysis of communication between users and their followers demonstrated that a substantial proportion of users follow other users from the same cultural groups as seen in Figure 7.4. This might indicate, that our users are more engaged in following others in their specific area of interest, in particular countries.

Next, we investigated whether the dimension or country location of user and follower match, as well as other user-related features having an influence on communications between Twitter friends. For this, we created a classification model based on decision tree and logistic regression techniques while using the aforementioned 16 features.

For creating the predictive model of user responses to their friends, we wanted to create a dataset with two categories of response levels regarding the level of "Interestingness" of user j to his/her follower i. Let r_{ij} be the frequency of i-th follower response (retweet or reply) to the user j. We can define the maximum and minimum levels of the follower i response as $\max_i r_{ij}$ and $\min_i r_{ij}$ respectively.

Next, we created the "Interestingness" function that returns one when user j gets the most of responses from the follower i; when the user gets the least of responses, its output is zero. We thus want to focus on the "extreme" cases when users have an interest or do not have an interest to their friends' content. Therefore, we disregard data instances with the "undefined" values of "Interestingness" while training our model with two categories regarding two levels of "Interestingness" calculated as follows:

$$Interestingness_{ij} = \begin{cases} 1, & \text{if } r_{ij} = \max\limits_i r_{ij} \text{ or } \min\limits_i r_{ij} = \max\limits_i r_{ij} \\ 0, & \text{if } r_{ij} = \min\limits_i r_{ij} \text{ and } \min\limits_i r_{ij} \neq \max\limits_i r_{ij} \\ undefined, & \text{otherwise} \end{cases} \tag{7.1}$$

Figure 7.5 shows that the follower response distribution is skewed towards a smaller number of follower responses. In a case when a user gets all responses of a follower (this value can be equal to one when the follower sent just one reply or retweet during our data collection), we still consider that the follower is interested in the followed user's content. Therefore we tested if follower's minimum response level equals to the maximum response level ($\min\limits_i r_{ij} = \max\limits_i r_{ij}$).

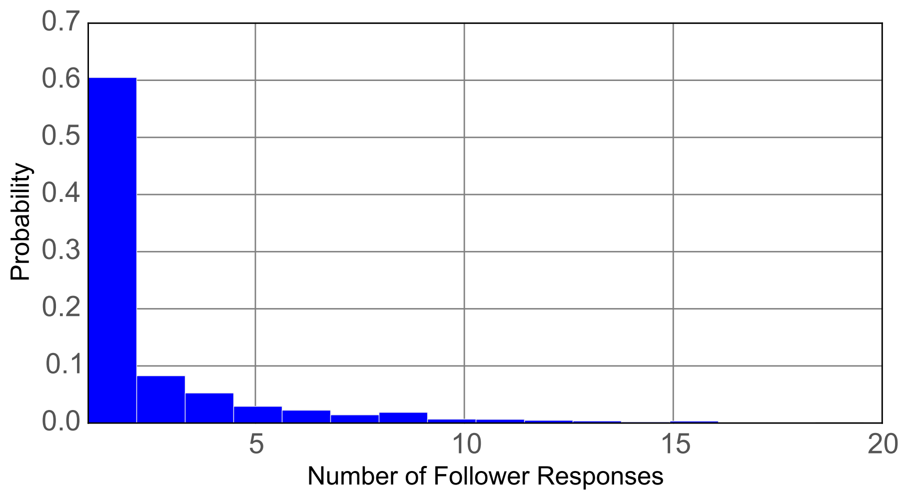


Figure 7.5: Follower responses to their friends
(based on 4574 user-follower response pairs, the response numbers were in the range [1..59], the average response number was 2.74 with std. deviation 3.65)

To find out what is important for predicting user responses to the friends of a user, we performed the following steps:

- Using formula (7.1), we assigned our initial Twitter users to two classes of the "highest interestingness" ($Interestingness_{ij} = 1$), and "lowest interestingness" ($Interestingness_{ij} = 0$) while disregarding the users with undefined values of the "Interestingness" score.
- It is important to mention, that of 4574 user-follower pairs, only 343 "Interestingness" values were of zero (no interest). Therefore, to train our model using two categories of equal size, we randomly select 343 instances of 1 value (highest "Interestingness).
- Next, having the "Interestingness" values representing a follower's interest (or "responsiveness") to his/her friend as labels, and user-related and follower-related features as features, we created our classification models.

Decision Trees. Using CountryMatch, DimMatch, FCMatch, FLangMatch, FTimezMatch, FTimezMatch, LangMatch, FCountries, FInfluence, FLanguages, FTimezones, Influence, Languages, Mentions, Mobility, Replying, Tagging, Timezones, and Weekends features for predicting 686 instances of "interestingness" with help of the decision trees classification technique. For evaluating our classification model, we split 686 records into training (75%) and testing (25%) sets. It achieved 65% of accuracy, 69% precision, 63% of recall and 65% of F1-measure. Next, we performed "leave-one-out" cross-validation while trained on all the ranks data except for one data point. When training on 685 instances, the classification accuracy was outperforming a random classifier in only about 15%. However, the created model enabled us to calculate variables' importance presented in Table 7.4 (Features Importance column).

Parameter	D. Tree	Logistic Regression Results, pseudo $R^2 \approx 0.23$, sensitivity ≈ 62, specificity ≈ 72							Marginal Effects					
	RFI	Odds Ratio	z	$P > \|z\|$	β	Std. Err.	95% Conf.Int.		$\frac{dy}{dx}$*	Std. Err.	z	$P > \|z\|$	95% Conf.Int.	
Intercept		172.7	2.61	0.01	5.15	1.97	1.29	9.02						
CountryMatch*	2.22	0.46	-1.48	0.14	-0.78	0.53	-1.82	0.25	-0.14	0.09	-1.49	0.14	-0.33	0.04
DimMatch*	3.91	4.45	2.31	0.02	1.49	0.64	0.23	2.76	0.23	0.11	2.35	0.02	0.04	0.49
FCMatch*	9.63	0.68	-1.96	0.05	-0.38	0.19	-0.76	-0.00	-0.07	0.03	-1.98	0.05	-0.14	-0.00
FLangMatch*	100	0.09	-10.03	0.00	-2.41	0.24	-2.88	-1.94	-0.43	0.03	-14.23	0.00	-0.49	-0.37
FTimezMatch*	7.20	1.20	0.86	0.39	0.18	0.21	-0.23	0.59	0.03	0.04	0.86	0.39	-0.04	0.11
LangMatch*	6.50	0.79	-0.67	0.50	-0.24	0.35	-0.93	0.46	-0.04	0.06	-0.67	0.50	-0.17	0.08
FCountries	24.79	1.16	1.44	0.15	0.15	0.11	-0.05	0.36	0.03	0.02	1.45	0.15	-0.01	0.06
FInfluence	74.49	0.26	-2.15	0.03	-1.36	0.63	-2.59	-0.12	-0.23	0.11	-2.17	0.03	-0.46	-0.02
FLanguages	56.84	0.85	-3.05	0.00	-0.16	0.05	-0.26	-0.06	-0.03	0.01	-3.12	0.00	-0.05	-0.01
FTimezones	25.78	0.98	-0.76	0.44	-0.02	0.03	-0.07	0.03	-0.00	0.01	-0.77	0.00	-0.01	0.01
Influence	70.61	0.94	-0.08	0.93	-0.06	0.73	-1.49	1.37	-0.01	0.13	-0.08	0.93	-0.27	0.25
Languages	23.37	0.78	-2.88	0.00	-0.24	0.08	-0.41	-0.08	-0.04	0.01	-2.94	0.00	-0.07	-0.01
Mentions	2.08	0.64	-0.62	0.53	-0.45	0.72	-1.86	0.96	-0.08	0.13	-0.62	0.53	-0.34	0.17
Mobility	2.96	0.14	-1.78	0.07	-1.97	1.10	-4.13	0.19	-0.35	0.20	-1.8	0.07	-0.74	0.03
Replying	28.60	1.08	0.05	0.96	0.08	1.59	-3.04	3.20	0.01	0.29	0.05	0.96	-0.55	0.58
Tagging	27.93	1.11	0.31	0.75	0.10	0.32	-0.53	0.73	0.2	0.06	0.31	0.75	-0.09	0.13
Timezones	10.82	0.69	-1.65	0.10	-0.37	0.22	-0.81	0.07	-0.07	0.04	-1.66	0.09	-0.14	0.01
Weekends	77.61	2.04	1.37	0.17	0.71	0.52	-0.30	1.73	0.13	0.09	1.38	0.17	-0.05	0.31

Table 7.4: Relative features importance in followers' response prediction test using decision trees, logistic regression (predicted logit of interest) and logit Marginal Effects, statistically significant (with $p < 0.05$) are shown in bold font. (*) The derivative $\frac{dy}{dx}$ (slope) shows the marginal effect values for the respective variables.

Surprisingly, relative importance statistics based on decision trees presented CountryMatch, DimMatch and LangMatch within the five least important features set. This is why we could not accept our Hypothesis H2.3.1, stating that user and follower's country locations and languages match are amongst the most important prediction parameters for the user responses. The most important features were FLangMatch, Weekend, FInfluence, Influence, and FLanguages. It seems that Country and Dimension match were not as important for the user response predictions, when using decision trees classification model. We explain this by the possible biases in our dataset towards the most active users, including also broadcasting agencies. To further assess variables' likelihood and effect on users' interest towards their friends, we perform logistic regression and compute their marginal effects.

Logistic Regression. We performed logistic regression analysis with the Statsmodel Package[4]. Table 7.4 presents logistic regression results considering binary dependent variable of user interest in user's friend coded as 0 (no responses) and 1 (most of responses). Some of the explanatory variables (marked with *) were categorical and coded as True or False when the compared values were matched or not. The overall model was statistically significant with log-likelihood ratio p-value less than 0.001. The pseudo R here cannot be interpreted as a measure of variance such as in the least squares regression.

The logistic regression showed the significance of the FLangMatch and FInfluence included in the top most important features derived with decision trees. Interestingly, DimMatch and Languages were also statistically significant. Marginal effects statistics presented in the last three columns in Table 7.4 showed that DimMatch is associated (statistically significant) with 23% higher probability of user response. FLangMatch, FInfluence and FLanguages are related to the statistically significant lower probability of replying in 43%, 23% and 3% respectively. However, Influence was statistically insignificant in our logistic regression model. This is why we could not strictly accept our Hypothesis H2.3.2.

Hypothesis Revisited and Discussion. One of the objectives of this analysis

[4]http://statsmodels.sourceforge.net/

was to investigate the possibility of exploiting microblogs to detect a country location of a user. For this, we considered several countries on Twitter, whose users were deemed most active in our sample set. Firstly, we observed that country-name metadata of Twitter users was matched with the country-name metadata of their followers in only 30% of our users. Therefore, we agreed with the H2.1 hypothesis stating that in most cases a country location extracted from a user and followers' metadata is insufficient for providing cues on user origins in our dataset (based on the data collected in one week). Secondly, we found out that the information publicly available in user metadata and followers' network enabled us to predict user country locations with a considerable accuracy of 90% or more for the best feature selection strategies we analysed while answering RQ2.1. The most successful feature combinations included both elements, Followers-related data FOLLOWERS, and User-related data LANGUAGE (RQ2.2).

However, usage of BEHAVIOUR together with LANGUAGE and FOLLOWERS feature sets did not provide any improvement. Therefore, we cannot strictly accept H2.2.2 without considering other features taking part in the classification strategy. Nevertheless, as a solution to address the need for user profiling to provide an improved user experience online, we might suggest exploiting user behavioural patterns or other well-performing features set combinations instead of directly asking user locations. This way, we could satisfy user preferences towards sharing content, times and ways of communicating with other users, whilst respecting privacy.

Overall, our results reveal a global orientation of our Twitter users in our dataset. Country Match and Language Match were not highly ranked (relative importance in decision trees), neither statistically significant features (logistic regression results). This is why we could not accept hypothesis H2.3.1. Dimension Match was more important than Country Match, and also statistically significant, leading to the improved probability of user replies. The importance of the cultural dimension match is specially seen when considering the previous study [79] of Twitter information flows (in respect to user mention statistics) prediction improvement when adding cultural dimensions of the Hofstede model outlined in [112]. Interestingly, user Influence was ranked in the top of feature importance in our decision tree results, while logistic regression showed no statistical significance, and FInfluence was even more important and significant. Thus, users with more influential followers might get fewer replies. When users have followers, with matching majority language, their reply probability decreases by 43%, which still supported the finding on the global nature of communication on Twitter. However, DimMatch should not be underestimated and requires a further investigation into further friend or related content recommendation experiments. As discussed in [142], cross-border user communication in Twitter is influenced by the geographic and language-specific proximity of involved Twitter users. Our findings, however, reveal culture-specific communication preferences in global communication in Twitter.

Privacy. Even though the country and user cultural dimension detection experiments open possibilities for adaptation, also concerns in regard to privacy issues could be raised, in particular for users with open profiles in Social Networks. Microblogging metadata, followers' network, and user-generated content enabled us to predict user country locations with considerable accuracy. Avoiding sharing location information in Twitter metadata might help in preserving user whereabouts only to a certain extent. Even language mined from user content or defined in the user profile provides insight into user locations to a certain extent. Therefore, for a privacy-concerned user, withdrawal from microblogging or closing open profiles is recommended.

Sampling Biases. It is important to mention that our data collection method is prone to sampling biases. Using Twitter sample, we might be biased towards the most active users such as event or news broadcasters, which requires further analysis.

On Sociological Models Usage. The Internet brings users from diverse cultures together, however, understanding their needs and requirements to realise quality features for web applications is challenging. Applying sociological models to assess website quality as perceived by users might not be trivial due to globalisation, since the new e-culture of individuals often does not comply with rules described in models referring to differences in cultural personality [223]. Therefore, we might require new approaches to study user behaviour online while respecting privacy.

Limitations. Our sample set contains users from the most active countries, providing geographical locations in metadata. The user-generated content was collected for one week. It is reasonable to assume that real-life activities might affect user behaviours. This is why we plan to explore user microblogging activities while assessing different models and feature combinations for predicting user origins and communication patterns to evaluate our approach in a larger time frame and extended locations set.

7.5 Conclusions

We analysed microblogging activities for persons from the top 13 most active countries on Twitter. We investigated different feature sets extracted from the microblogging content and metadata of publicly available tweets. Our findings reveal that combining user-related microblogging features and features obtained from a followers' network enables user country prediction with an accuracy of more than 90%. Considering sociological studies and previous works on behavioural differences online, we proposed an approach for mining individual culture-specific microblogging preferences which we abstracted from country information, often revealed in user metadata and content. This provides insights on the adaptation for web applications' and personalisation options to preserve individual privacy while improving user satisfaction online by providing application features/content which is of interest for the cultural origins of the user. Finally, we investigated user interactions and found out that users from the same cultural groups tend to communicate more with each other. Moreover, user communication amongst cultural groups in the long term could be analysed, aiming to uncover recommendation approaches for further improving user experience on the Web. Also, other Twitter-specific features such as tweeting frequency, topics found in the tweets' content, and an in-depth tagging behaviour could be explored. Next, we proceed with uncovering privacy preferences in Twitter microblogs.

Chapter 8

Privacy Settings Usage in Twitter

> "Friends don't spy; true friendship is about privacy, too."
>
> - Stephen King, Hearts in Atlantis [133]

In this chapter, we analyse Twitter settings usage and their relation to the usage of main Twitter features. The main findings of this chapter were previously published in the workshop paper "Usage and Consequences of Privacy Settings in Microblogs" [53]. Herein we provide more details on the experimental setup and results. We conclude with a discussion and our insights into implications for further research.

Social networks can reveal much information about the personal life of users. This information may be explicitly provided or harvested with machine learning tools, to mine specific user traits, for instance, such as home locations [156], out of microblogging data. It might be argued, that the primary purpose of microblogging is to share information and network, such that microblogging content should be openly available for others. However, human privacy needs should be respected and better supported online and in research as well.

To address this controversy, we investigate how microbloggers exploit privacy and geolocation setting controls in [53]. This chapter is built on findings published in [53] and further discussed data mining approaches applicable to microblogs and related ethical considerations. Additionally, we analyse how the persons from the defined cultural groups exploit privacy setting controls in Twitter.

Feature	Description	Feature	Description
INFLUENCE	Ratio of Followers to Following (Friends)	STATUSES	Number of published tweets
FAVOURITES	Number of favourites user posted	LISTED	Number of lists in which user was included
FOLLOWERS	Number of followers	FRIENDS	Number of friends
SOURCES	Number of Twitter applications that were used to post	CHANGES	Number of privacy setting (enabling/disabling geolocation services or protecting/opening user profiles) changes

Table 8.1: Features analysed in respect to privacy settings

8.1 Research Questions

In this chapter, we address the research Objective 3: "Monitoring Twitter privacy settings usage". Firstly, we want to find out whether different cultural groups and countries have different fractions of protected and open profiles. For answering the RQ 3.1. "Are there differences in privacy settings usage by different cultural groups?", we setup up the following hypotheses:

- H3.1.1. MA users have a larger fraction of open profiles compared to RE and LA users.
- H3.1.2. RE users have a smaller fraction of open profiles compared to MA and LA users.

The next research question RQ 3.2. of the Objective 3 was to find out if protecting user accounts hampers an effective communication in Twitter. We selected a non-exhaustive list of features mostly available in the Twitter profile and listed in Table 8.1. Next, we compared protected users activity with microblogging activity of users with open profiles, and tested the respective statistical hypothesis as follows (we based our hypothesis on the assumption that users with protected profiles are less active on Twitter, not so influential and have smaller social networks size when compared with users with open profiles, which content is openly available, can be found with search engines and potential followers do not require an approval for establishing a following relationship [1]):

- H3.2.1. Number of friends (FRIENDS feature) which user follows is fewer for protected user accounts;
- H3.2.2. Number of followers (FOLLOWERS) is smaller for the protected user;
- H3.2.3. User influence (INFLUENCE), defined as the ratio of followers to friends, is smaller for the protected users;
- H3.2.4. Status updates (STATUSES), or twitter microblogs posted by the user, are less posted by protected users;
- H3.2.5. Number of lists (LISTED) in which user was included is smaller for the protected users;
- H3.2.6. Number of favourites (FAVOURITES) is smaller for protected users.

Since the privacy control designs differ amongst Twitter software clients, we also analyse the use of Twitter client software. Our last hypotheses assume that protected users tend to exploit and try out different software products and setting changes at the beginning of Twitter usage, to find out which software and settings fit the best to their needs:

- H3.2.7. Number of setting changes (CHANGES) of protected users is greater compared with the "open users" in average.
- H3.2.8. Number of software clients (SOURCES) of protected users is greater compared with the "open users".

Finally, we discuss the security and social implications of culture-related privacy preferences in microblogging (RQ 3.3).

[1] see "About public and protected Tweets" at https://help.twitter.com/en/safety-and-security/public-and-protected-tweets

8.2 Methodology

Despite the “openness” culture widely appreciated in Twitter, we argue that for some of the users, privacy still matters. However, due to the lack of information on privacy issues and insufficiency of the privacy-protecting mechanisms online, it is paramount to investigate real user needs, which further might be governed by different purposes of and modes of microblogging usage.

To approach the problem of microbloggers’ privacy online, we overview the privacy protection means provided by Twitter. We do not provide or suggest privacy-protecting solutions, which are in practice implemented by SN and mobile app developers. Instead, we analyse user behaviour online to raise awareness towards protecting privacy on Twitter. We observe how Twitter users exploit twitter privacy settings. Particularly, we focus our analysis on the usage of Twitter profile protection and geographic location sharing features. For this, with the help of Twitter API, we followed a set of users for about six months and analysed their privacy controls’ usage. We distinguish between different usage purposes, which we relate to the use of geolocation and profile protecting features. Our main contribution is to study the usage of online privacy controls for a set of selected users in Twitter.

8.2.1 Experimental Setup

It is important to mention that users change their settings over time and we analyse which settings are mostly exploited by each particular user. This is why we follow users starting their microblogs on the same day and observe their setting changes for about half of year as follows:

- STEP 1: Collect a set of 21,600 users (we planned to visit each of these profiles on a daily basis considering Twitter rate limitations of 350 requests per hour for an authenticated user, with the help of three user accounts; see 5.3.2 for more details on Twitter access) registered with Twitter on 26th November 2014 by listening to the Twitter sample stream for about three days, from 26th to 29th November 2014 [2];

[2]Please refer to Table G.1 in appendices

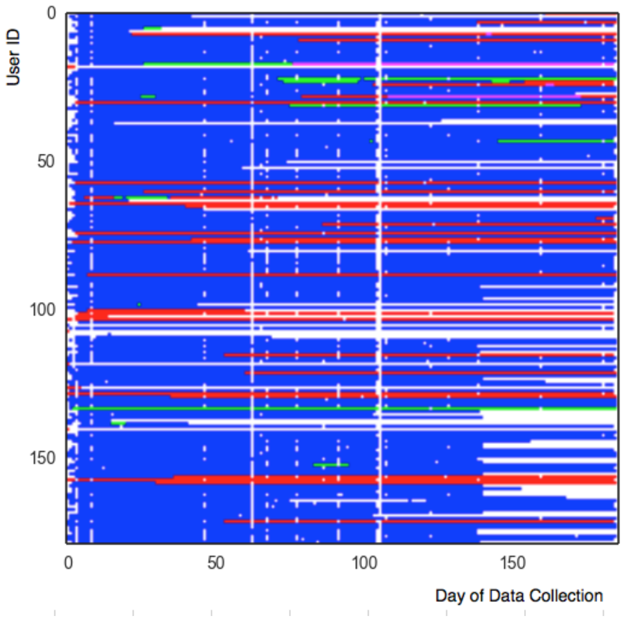


Figure 8.1: Dynamics of Twitter profile changes (based on the “Privacy” dataset, slice of 195 randomly selected users and 186 days of data collection; blue points denote open user profiles without geolocation, red points for open profiles with geolocation, green for protected profiles without geolocation, magenta points for protected profiles with geolocation enabled)

- STEP 2: Visit the selected user profiles in about six months and monitor their usage of the geographic location sharing and profile protection features, and statistics on status updates and friendship/followers network growth in time;

- STEP 3: Analyse and interpret the collected data for addressing the hypothesis stated above.

8.2.2 Data Collection

In the first step, we collected a sample of tweets from 26th to 30th November 2014, from which we randomly selected 21,600 users (considering Twitter rate limitations and our data collection setup), which are registered with Twitter on 26th of November 2014 to be further followed for the next six months. We exploited our three Twitter accounts with Twitter Representational State Transfer (REST) Application Profile Interface (API) for visiting user profiles of the selected users. These user profile settings were stored into the “Privacy” dataset (a summary is in the

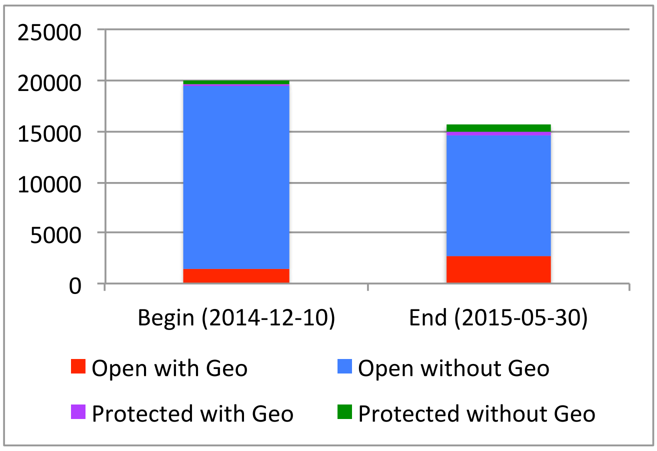


Figure 8.2: Settings changes in six months

Table D.5 [3]) including information on the number of published tweets, followers, friends, listed, favourites count; and also on the geolocation settings turned on, verified (if a person is a celebrity or verified on Twitter) and protected (the protected accounts' content is only visible to the user friends) fields' settings.

We experienced some interruptions while storing user profile settings as white vertical strokes appear in Figure 8.1. This picture represents the exploited privacy setting per user per day since we visited each user profile only once a day. It is important to mention that we do not consider daily change patterns due to Twitter REST API limitations[4] imposed. It is seen from the colour differences, that some of the users change their settings to enable geolocation sharing (change from blue colour to red), while others protect their profiles (green colour).

We also observed an increase of protected profiles and geolocation sharing profiles towards the end of data collection as seen in Figure 8.2.

Having the data on Twitter usage and user settings for the sampled half of year, we aggregated the most exploited settings of client software usage, geolocation sharing and profile protection, and the maximal number of tweets, friendships, lists and favourites. These data comprised a "Privacy Aggregated" data table (see Table D.5 summarising the data), which we further study to find out general behaviour differences amongst Twitter users having certain privacy and geolocation sharing setting preferences.

[3]It is important to mention that not all of 21,600 user profiles ended up in our dataset. We could explain this that some of the accounts were blocked by Twitter or closed.

[4]https://dev.twitter.com/rest/public/rate-limits

Table 8.2: Privacy and geolocation settings in "Privacy Aggregated" table:
T (True) for enabled setting, otherwise F (False)

Geolocation	Protected	Verified	Description	Number of Users
Open User Profiles				
F	F	F	Open profiles without geolocation setting activated	17182
T	F	F	Open profiles with geolocation setting enabled	2973
Protected User Profiles				
F	T	F	Protected profiles without geolocation	733
T	T	F	Protected profiles with geolocation setting activated	243
Disregarded User Profiles				
F	F	T	Protected profiles of public persons (or celebrities) disabled geolocation setting	2

8.3 Findings

8.3.1 Setting Groups

To summarise, about 95% of our users keep their user profiles openly available. Geographical location sharing feature is exploited mostly by 15% of our users. Table 8.2 lists our main setting combinations. It is important to mention that we did not have any other public user settings than "FFT", referring to geo-disabled and open profiles. The greatest group of 17182 users prefer open profiles without geographic information available. The second largest group of 2973 users have public profiles with geolocation setting enabled.

We are interested to find out whether there are any differences in Twitter usage for users with protected and open Twitter profiles. When concluding with our hypothesis outcomes, we consider the overall open and closed profiles without respect to the geo-location sharing. We are however aware that there might be differences in respect to different modes of Twitter usage when users also share or do not share their geographic locations. This is why we performed additional statistical tests while breaking down the user groups for open and closed user profiles into separate groups for geo-enabled and geo-disabled profiles. This enabled us to do more-finer grained conclusions (differences between geo-enabled and geo-disabled user profiles) while addressing the main hypothesis stated in section 8.1 and referring only to open or protected user profiles (page 139).

8.3.2 Twitter Feature Usage Compared

Having "Privacy Aggregated" table (see Table D.5 in appendices) with user preferences accumulated during half of the year of data collection, we compared networking and other Twitter features usage, as shown in Figures 8.3 and 8.4. In the box plots, shown averages also include "verified" users (FFT) of two public persons, in our case a music band and politician, which are disregarded in the group comparisons when performing unequal variances statistical tests (next section, Table 8.3). We might see that these two users have larger social networks when compared with other user groups. They also included into more listings as shown in Figure 8.4.

Friends and Followers. Interestingly, the cumulative distribution function with logarithmic scale helps to visualize differences between user groups. Particularly, protected users with enabled geolocation services (TTF) have a greater number of FRIENDS and FOLLOWERS compared with users with closed profiles without geolocation settings (FTF). The INFLUENCE of the last mentioned is however slightly larger than of the former ones.

Status updates, Listed and Favourite Counts. Users (TTF and TFF) with protected and open profiles, with geolocation services enabled do post more STATUSES on average. TTF and FFF has no significant differences in their status updates (Table 8.3). Users with open and geo-enabled profiles (TFF) tend to be included into more lists (LISTED), while the protected users with geo-sharing settings (TTF) compete in the number of FAVOURITES with open users who do share their geolocations (TFF) as well.

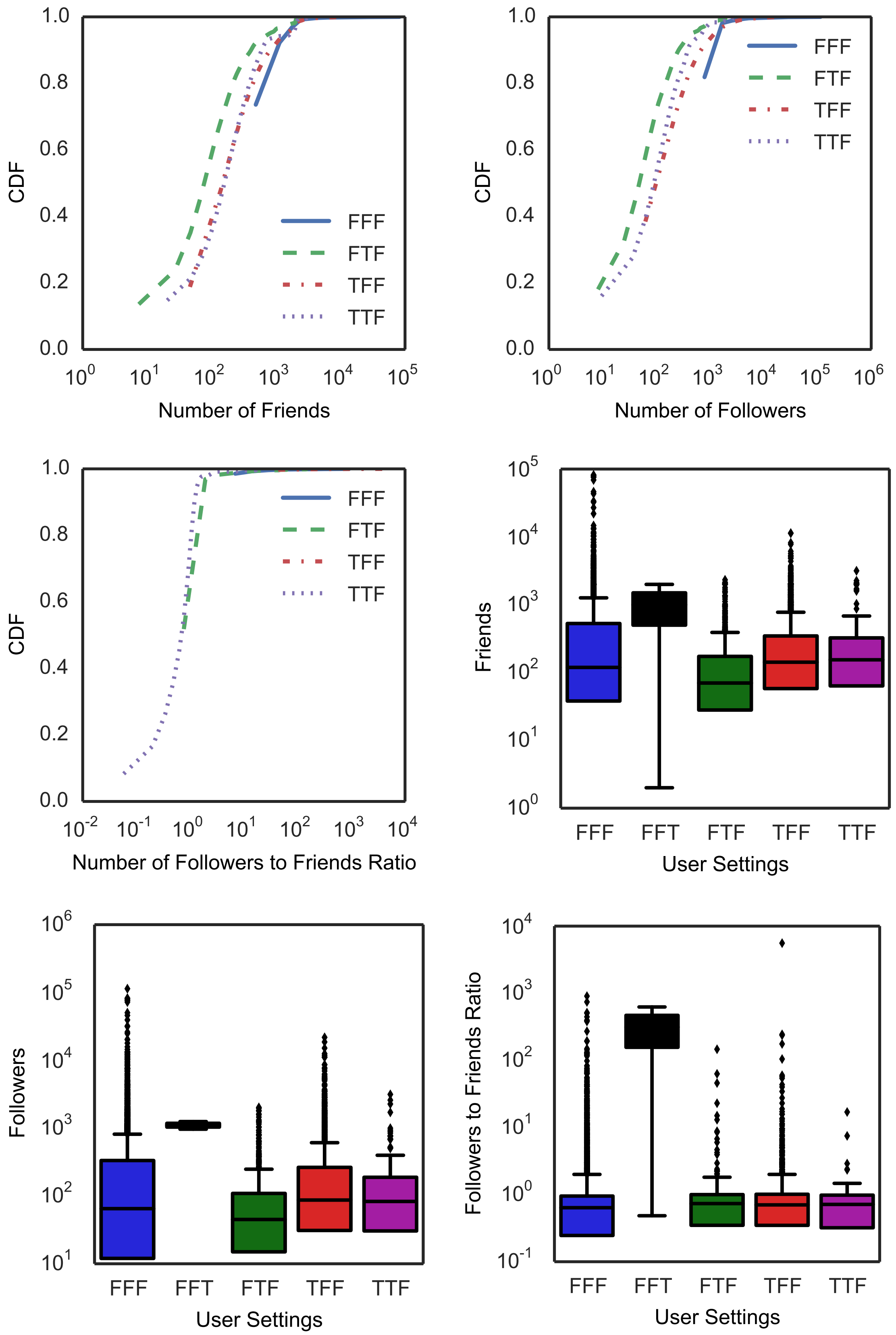


Figure 8.3: Networking with different profile settings

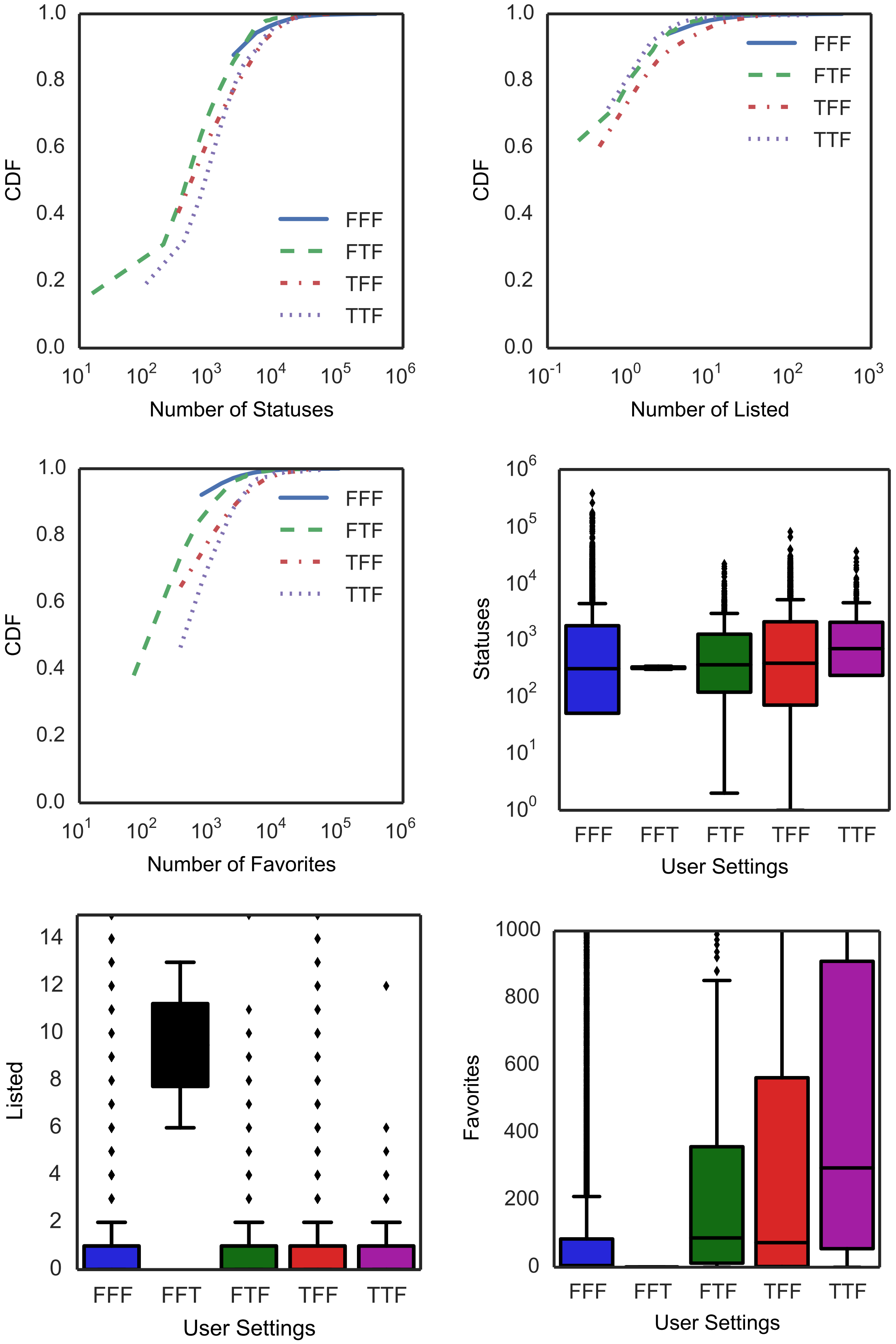


Figure 8.4: Features usage with different profile settings

Feature	μ_{gr_1}	σ_{gr_1}	μ_{gr_2}	σ_{gr_1}	t	p
$Group_1$ ($N_{gr_1} = 17182$): FFF vs. $Group_2$ ($N_{gr_2} = 733$): FTF						
INFLUENCE	2.72	105.46	1.76	19.31	0.88	0.37
STATUSES	1810.87	7844.50	1208.69	2321.30	5.75	< 0.01
FAVOURITES	323.78	1793.38	503.16	1782.19	-2.66	< 0.01
LISTED	1.27	7.37	0.74	2.53	4.85	< 0.01
FOLLOWERS	367.30	1902.00	119.79	245.46	14.46	< 0.01
FRIENDS	421.12	1591.38	183.78	349.27	13.39	< 0.01
SOURCES	1.03	0.20	1.03	0.18	0.26	0.78
CHANGES	1.08	0.32	1.92	0.39	-56.34	< 0.01
$Group_1$ ($N_{gr_1} = 17182$): FFF vs. $Group_2$ ($N_{gr_2} = 243$): TTF						
INFLUENCE	2.72	105.46	0.75	1.18	2.43	< 0.05
STATUSES	1810.87	7844.50	2157.59	4259.59	-1.23	0.21
FAVOURITES	323.78	1793.38	1301.08	4150.01	-3.66	< 0.01
LISTED	1.27	7.37	1.27	10.93	0.00	0.99
FOLLOWERS	367.30	1902.00	182.81	349.50	6.90	< 0.01
FRIENDS	421.12	1591.38	298.31	474.33	3.74	< 0.01
SOURCES	1.03	0.20	1.03	0.17	0.26	0.79
CHANGES	1.08	0.32	2.76	0.74	-35.13	< 0.01
$Group_1$ ($N_{gr_1} = 2973$): TFF vs. $Group_2$ ($N_{gr_2} = 733$): FTF						
INFLUENCE	2.39	68.48	1.76	19.31	0.43	0.66
STATUSES	2144.56	4579.74	1208.69	2321.30	7.79	< 0.01
FAVOURITES	1128.26	3391.42	503.16	1782.19	6.90	< 0.01
LISTED	1.81	6.14	0.74	2.53	7.29	< 0.01
FOLLOWERS	320.75	978.62	119.79	245.46	9.99	< 0.01
FRIENDS	339.87	632.67	183.78	349.27	8.99	< 0.01
SOURCES	1.05	0.23	1.03	0.18	2.18	< 0.05
CHANGES	1.68	0.69	1.92	0.39	-12.60	< 0.01
$Group_1$ ($N_{gr_1} = 2973$): TFF vs. $Group_2$ ($N_{gr_2} = 243$): TTF						
INFLUENCE	2.39	68.48	0.75	1.18	1.30	0.19
STATUSES	2144.56	4579.74	2157.59	4259.59	-0.04	0.96
FAVOURITES	1128.26	3391.42	1301.08	4150.01	-0.63	0.52
LISTED	1.81	6.14	1.27	10.93	0.76	0.44
FOLLOWERS	320.75	978.62	182.81	349.50	4.80	< 0.01
FRIENDS	339.87	632.67	298.31	474.33	1.27	0.20
SOURCES	1.05	0.23	1.03	0.17	1.51	0.13
CHANGES	1.68	0.69	2.76	0.74	-21.98	< 0.01
$Group_1$ ($N_{gr_1} = 20155$): Open FFF and TFF vs. $Group_2$ ($N_{gr_2} = 976$): Protected FTF and TTF						
INFLUENCE	2.67	100.86	1.51	16.75	1.30	0.19
STATUSES	1860.09	7454.24	1444.94	2952.54	3.83	< 0.01
FAVOURITES	442.45	2125.85	701.82	2603.56	-3.06	< 0.01
LISTED	1.35	7.20	0.87	5.87	2.45	< 0.05
FOLLOWERS	360.44	1795.96	135.48	276.22	14.57	< 0.01
FRIENDS	409.13	1489.56	212.30	387.15	12.12	< 0.01
SOURCES	1.03	0.21	1.03	0.18	0.74	0.45
CHANGES	1.17	0.45	2.13	0.62	-47.51	< 0.01

Table 8.3: Unequal variances t-test for various Settings: Mean (μ), Standard Deviation (σ), Welch's Test statistic (t), two-tailed p-value (p). The first four comparisons were performed additionally to extend the last comparison of open and protected user profiles, which is used for concluding on RQ3.2.

Client	FFF	TFF	FTF	TTF
Twitter for Android	24%	52%	38%	51%
Twitter for IPhone	23%	26%	40%	34%
Twitter Web Client	14%	13%	9%	6%

Table 8.4: Top three Twitter Clients: Software (percent of users preferring it, overall we observed 501 different Twitter client names in our dataset including these three most used)

Frequency of Setting Changes and Software Clients. Table 8.3 shows statistically significant (p-value < 0.01) differences in CHANGES feature between open and closed user profiles. Geolocation sharing users (TFF) changed their applications (SOURCES) more often compared with the users with protected accounts (FTF), which in turn change their settings more on average. Interestingly, users with protected profiles (FTF) change their settings more than geolocation sharing users (TFF), which in contrast use more SOURCES. Despite user profile protection, Twitter still provides information on client software usage shown in Table 8.4 listing top three Twitter clients. Twitter for Android software is used by about 30% of our users, followed by Twitter for iPhone used by 24% and Twitter Web Client used by almost 14% of users. Twitter Web Client is used the least by geo-enabled users with closed profiles (TTF). It seems that this user group prefers to use mobile device applications instead.

8.3.3 Cultural Differences in Privacy Settings Usage

As seen in Figure 8.5, MA users have a larger fraction of open profiles, while RE have the smallest fraction of open profiles. Therefore, we could accept our hypotheses H3.1.1 and H3.1.2. It seems that privacy perceptions or needs differ amongst the cultural groups analysed.

8.3.4 Facebook References

Since we collected user tweets for the first days of data collection, we were able to capture also 60 users sharing their Facebook pages in tweets. We were interested in their purpose of Facebook usage. We manually labelled these 60 users into the 5

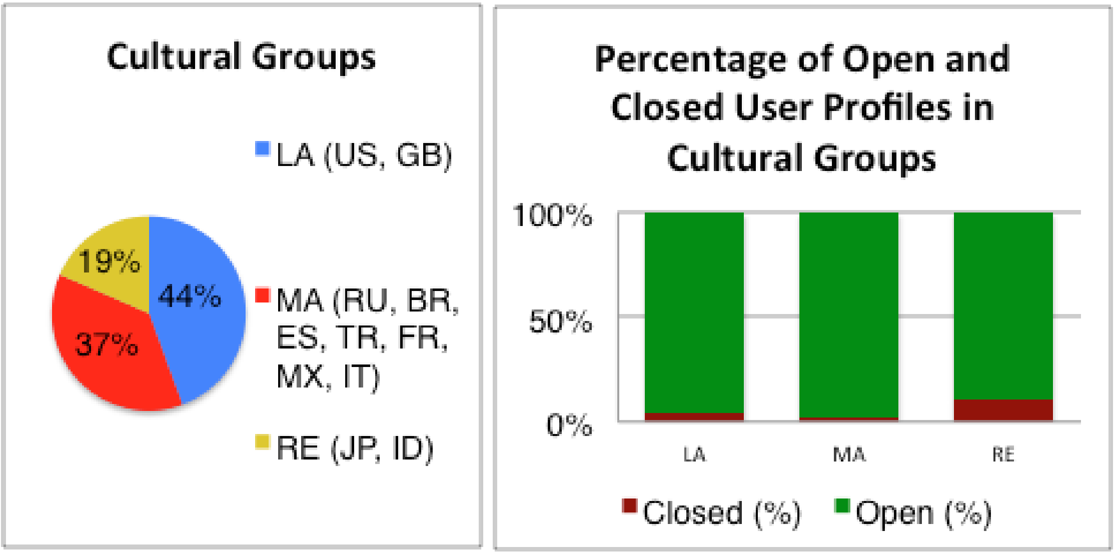


Figure 8.5: Privacy settings' preferences and cultures

main categories including "Personal" (personal pages), "Business" (including companies such as news broadcaster, Instagram, farm, a t-shirts seller amongst others), "Music" (including music bands, an entertainer and a professional pianist), "Community" (strokes support and a sport team), and also two categories "Not Available" and "Not Accessible", in which two user pages become respectively closed and not available anymore. Table 8.5 below shows the breakdown of users by categories and defined in the Facebook gender.

Next, we compare Twitter and Facebook usage purposes for these 60 users. We found out that Twitter users with Facebook profiles were classified into "Music" category are the most influential in respect to Followers to Friends ratio. However, Business and Community users have higher inclusions in Twitter lists. This might explain why users associated with Personal and Music Facebook accounts have highest favouring averages compared with the rest. It seems that favouring Twitter feature might be used to seek attention from other users as we mentioned in respect of Twitter users with the protected accounts.

		Gender (in percentage)			Averages		
Category	Users #	Female	Male	Not Defined	Influence	Favourites	Lists
Personal	36	94.7	94.7	–	1.3	27.5	0.03
Music	9	5.3	–	36.4	2.1	13.1	0.00
Business	8	–	–	36.4	0.5	4.4	0.12
Community	5	–	–	22.7	0.3	6.8	0.2
Not Accessible	1	–	–	4.5	0.4	7.0	0.00
Not Available	1	–	5.3	–	1	5	0.00

Table 8.5: General statistics of 60 Users from our Twitter dataset shared their Facebook account

	Hypothesis	Conclusion
H3.1.1	MA users prefer open profiles the most	Accepted
H3.1.2	RE users prefer closed profiles the most	Accepted
H3.2.1	Protected users have fewer friends	Accepted
H3.2.2	Protected users have fewer followers	Accepted
H3.2.3	Protected users are less “influential”	Rejected
H3.2.4	Protected users have fewer status updates	Accepted
H3.2.5	Protected users are less listed	Accepted
H3.2.6	Protected users have less favourites	Rejected
H3.2.7	Number of setting changes of protected users is greater compared with the open user	Accepted
H3.2.8	Number of software products (SOURCES) of protected users is greater compared with the open users	Rejected

Table 8.6: Hypothesis revisited: while comparing open (FFF+TFF) and protected (FTF+TTF) user profiles

8.3.5 Open vs. Closed (Protected) Profiles

Table D.5 shows the descriptive statistics of feature values, which are continous. Our sample distributions could not satisfy normality and equal variances assumptions in the majority of cases. This is why we selected Welch’s unequal variances statistical test (using Scipy Python library) to compare users with particular setting groups. This test allowed us to test if two independent samples have similar averages. When p-values were $p < 0.05$ we disregarded the null hypothesis that the averages are similar. Table 8.3 (page 148) shows results for comparing means between paired setting groups, which we used to accept or reject our research hypothesis revisited in Table 8.6. We arranged the table to compare settings groups for open profiles, including FFF (open profiles without geolocation services enabled) and TFF (open profiles with geolocation enabled), and protected profiles, including FTF (closed profiles) and TTF (closed profiles with geolocation services enabled). At the bottom of the table we placed the features comparison for answering our hypothesis statements for combined user groups, with open (FFF and TFF) and protected (FTF and TTF) profiles.

Overall, we accept H3.2.4 and H3.2.5 since users with publicly open profiles exploit the Twitter features STATUSES and LISTED the most compared with users with protected user profiles, with significance $p - value < 0.01$ for STATUSES ($\mu_{gr1} = 1860.09$ versus $\mu_{gr2} = 1444.94$) and $p-value < 0.05$ for LISTED ($\mu_{gr1} = 1.35$ versus $\mu_{gr2} = 0.87$) respectively. Generally, open profiles attract more followers ($\mu_{gr1} = 360.44$ versus $\mu_{gr2} = 135.48$) and have more friends ($\mu_{gr1} = 409.13$ versus $\mu_{gr2} = 212.30$) when compared with closed profiles (with significance $p < 0.01$ for both features), therefore we could accept our hypotheses H3.2.1 and H3.2.2. However, we could not find significant differences in the INFLUENCE feature for the geolocation enabled users (TFF and TTF). As seen from the bottom of the table Table 8.3, the INFLUENCE does not differ significantly (p-value = 0.19) for the users with open ($\mu_{gr1} = 2.67$) and protected profiles ($\mu_{gr2} = 1.51$). Similarly, number of SOURCES is comparable for both user groups ($\mu_{gr1} = \mu_{gr2} = 1.03$ with $p - value = 0.45$). This is why we cannot accept hypotheses H3.2.3 and H3.2.8.

Interestingly, users with protected profiles (FTF and TTF) tend to exploit FAVOURITES feature more actively in contrast with open profile users (FFF). However, when users with open profiles enabled their geolocation settings, their favouring statistics were not significantly different compared to the users with closed profiles (TTF). When comparing overall PROTECTED and OPEN user groups, we found that PROTECTED user profiles exploit significantly (p-value < 0.01) more FAVOURITES ($\mu_{gr2} = 701.82$) compared to OPEN user profiles ($\mu_{gr1} = 442.45$), therefore we reject H3.2.6.

Interestingly, the protected user profiles with geolocation feature enabled (TTF) showed no significant differences when compared with open user profiles (FFF and TFF) in the number of STATUSES, LISTED and SOURCES. Thus, these users quite active in their publishing behaviour, they included in lists and exploited various devices and software clients. It seems that we still need to have profile protecting feature to address this user group needs. Their motivations and purpose of microblogging usage could further be investigated with a user feedback.

Moreover, we found different cultural preferences towards protecting and opening user accounts in Twitter. Users from MA countries are more likely to prefer opened user accounts, while RE users prefer closing their accounts on average.

8.4 Discussion

8.4.1 Reflection on Our Results

Even though protecting Twitter profiles might seem to be counterproductive due to the microblogging nature of networking, borderless communication and word of mouth advertising. However, we noticed that 4.62% of our users prefer to close their accounts form the public view. We observed a two-fold increase in the number of protected accounts in a half year period of following user accounts. A word of caution should be said that protecting user accounts in microblogs could be misleading users into false perceived safety since personal data could still be automatically mined or revealed by online friends.

We found no significant differences in user influence and number of sources for protected and open user profiles. However, the number of status updates, listed, followers and friends were greater for the open user profiles. The open user profiles with geolocation enabled were the most active user group in terms of all features we analysed. One of the interesting findings is that protected user profile tend to favourite the most. Does it mean that they like to keep the favourite tweet for later and do not want to further propagate the tweet as when using retweet? Alternatively, favouring might mean that protected users personally appreciate authors of their favourite tweets and might motivate their following behaviour. Additionally, we could not find significant differences of posting status updates and number of list inclusions between geolocation enabled users with protected profiles and users with open profiles. We think that a further investigation into purposes of different microblogging usage modes and privacy preferences should be further investigated preferable with the feedback of the microbloggers actively using Twitter services.

8.4.2 Ethical Considerations and Twitter Research

To summarise, social applications and web services enable personal data collection in order to provide state-of-the-art communication tools. Providing users with complete control over their shared data is essential. In accord with [178], personal data gathering should adhere to national regulations and be performed with permission of the persons involved. The goal of data collection should be clearly communicated, while data collected should only serve the intended goal. The personal data should be stored, accessed, processed, and exploited securely to preserve human privacy rights [178].

Furthermore, openly available Twitter content and metadata could provide scientists with much required data for performing research experiments. However, would it be ethical to access user data without appropriate consent? The Twitter corporation provided their public tweets archive to the Library of Congress [224, 239]. Twitter states that tweets collection opens new perspectives for research and ways to retrieve information related to past events [233]. Twitter together with Facebook and Buzz also made their public content searchable via the Google Search engine [233]. The ethical dilemma of using Twitter data in research while protecting user privacy was discussed in [201] suggesting the main steps of supporting ethical research on Twitter:

- Experimental setup should be thoroughly described and communicated;
- User context should be considered while protecting privacy rights;
- User geographic locations should be well secured and scaled to the larger area to avoid compromising precise user locations;
- Researchers should avoid the use of other external resources for further inferring user details;

- When research studies microblogging behaviour of particular users rather than dealing with the aggregated statistics, Institutional Review Board's approval procedures should take place;
- Researchers should consider sharing settings defined in the user profile.

In our research, we preprocess the collected data to further work with aggregated user profiles while analysing user behaviour patterns and preferences. We thus avoid data retention of individual users, and anonymise usernames and tweets when needed.

Moreover, Twitter's terms of service postulate that it is users' responsibility to share their content online [245]. It states that users' content is published following the royalty-free licence and will be posted as provided by the user [245]. Twitter also informs that they do not disclose personally identifying information unless the user decides to publicly share it online or in special cases as defined in the privacy policy [245, 248]. Users also have control over the advertisement shown and the collection of browser information by Twitter [249].

8.4.3 Open vs. Commercial Exploitation

Commercial companies can also use Twitter as a marketing and communication tool for influencing their customers, as was studied in [30]. The effectiveness of Twitter and Facebook usage for hotel marketing purposes was investigated in [147] showing a strong relationship between social media user experience, hotel perceptions and further booking intentions. In opposite of the openly available social media content, followers' control could open a possibility to paid services [57]. However, this approach could change microblogging as it is now, and its impact on the openness of content and society at large requires further consideration.

8.4.4 Usage Purposes and Protected Messages

In contrast to the Twitter usage motivations discussed in [126], we do not aim to perform manual analysis of Twitter microblogs to determine the primary user motivation during microblogging. We also do not exploit the suggested classification of Twitter messages into four categories in accordance with the goal of microblogging. We of the opinion that there might be more Twitting purposes, such as organising content with the help of hashtags, or propagating the news with the help of retweets. There might be other user intentions including not only personal goals, but also marketing, advertisement and other goals of businesses and non-profit organisations. Besides good usage intentions, there also the antisocial behaviour of Twitter usage studied in [209], terrorism decision making using social networks [179].

As we observed in the previous section, there are users such as music bands and sports groups, which intentions might differ quite a bit from personal users often having a lesser influence and authority in their communication networks. We thus might partially agree with the user roles including information source and seeker described in [126]. The same user, however, might have different user roles in different communities [126].

Roughly we classify Twitter usage into personal and business/non-profit categories. Twitter is widely used by corporations, agencies and in show business to promote and advertise products or services. Twitter is a great messaging and news sharing platform, also providing business feedback from their customers. Sensitive corporate data can be protected and Twitter could provide a relatively secure messaging tool. All Twitter users can also block unwanted users and spammers, which eventually get suspended from using Twitter.

Personal Twitter accounts allow informal communication and networking for sharing personal messages, which is similar to Facebook status updates. Facebook allows building social networks based on real-life connections such as school classmates, friends and family members and enabling more sophisticated privacy controls. In opposite, Twitter provides a simple protection feature for blocking non-followers from seeing messages of a user. In short, we need to ask permission to follow protected user tweets. It seems sufficient for protecting sensitive and personal data, however, as we discussed above, the protected data still can be leaked from the user friendship network.

8.5 Conclusion

In this chapter, we analysed usage of Twitter profile protection and geolocation sharing controls. Since the majority of users prefer to keep their profiles open and avoid using geolocation sharing, we were interested to find out if users with protected profiles do not exploit Twitter features to their full advantage. For this, we statistically compared user groups of protected and openly available profiles regarding their status updates, contact networks and other features. We found out that protected users have smaller social networks, however, not less influential. Protected users actively favourite other content compared to the open-profile users without geolocation services enabled. When with users with protected profiles enable their geolocation services, their tweeting behaviour does not differ significantly from users with open profiles but without geolocation enabled. When users with open accounts enable geolocation services, they become the most active while using the aforementioned Twitter features except of privacy setting changes. Protected users change their software settings and prefer to use mobile Twitter client applications the most. It seems that users preferring to exploit protected profiles have their motivation to microblog and thus have their own privacy needs. However, despite of the small fraction of the protected user profiles, human privacy in microblogs cannot be underestimated, in practical applications and also in research. We suggest more thorough exploitation of the user-generated content while rising user attention towards possible privacy threats in microblogs. Social networks and data mining techniques can help in attaining user sensitive information despite of the privacy controls available.

Additionally, we observed cultural privacy preferences in microblogs while extending our experimental setup with our further developed user location detection classifier. Therefore we might suggest adapting default user privacy settings in accord with user cultural preferences in web applications and SN websites.

Moreover, despite the open availability of royalty-free users' data on Twitter and its value for research, users privacy should be duly supported in research. It is nevertheless paramount to respect privacy settings and seek IRB approval when dealing with individual user profiles. There are no standard guidelines for using the openly available content in research yet. However, it is clear that researchers should adhere to the rules imposed by Twitter for accessing and sharing their datasets while being accurate when disseminating potentially sensitive personal data. Sharing individual user geographic locations might lead to security harms. Other personal data collected from microblogs could also become sensitive when viewed in a particular context or provided with additional data from external resources. Therefore, we aim to exclude precise geographic information of users in our dataset comprised of publicly available tweets collected and maintained in accord with Twitter rules of conduct. For sharing our research datasets, we store only aggregated usage statistics. The initial tweets' IDs could be further communicated upon request when the related tweets are publicly available in the Library or other archiving services.

MINING MICROBLOGS FOR CULTURE-AWARENESS IN WEB ADAPTATION

Volume 2 of 2

Elena Alexandrovna Daehnhardt

Submitted for the degree of
Doctor of Philosophy

Heriot-Watt University
School of Mathematical and Computer Sciences
Department of Computer Science

August 2018

Abstract

Prior studies in sociology and human-computer interaction indicate that persons from different countries and cultural origins tend to have their preferences in real-life communication and the usage of web and social media applications. With Twitter data, statistical and machine learning tools, this study advances our understanding of microblogging in respect of cultural differences and demonstrates possible solutions of inferring and exploiting cultural origins for building adaptive web applications. Our findings reveal statistically significant differences in Twitter feature usage in respect of geographic locations of users. These differences in microblogger behaviour and user language defined in user profiles enabled us to infer user country origins with an accuracy of more than 90%. Other user origin predictive solutions we proposed do not require other data sources and human involvement for training the models, enabling the high accuracy of user country inference when exploiting information extracted from a user followers' network, or with data derived from Twitter profiles. With origin predictive models, we analysed communication and privacy preferences and built a culture-aware recommender system. Our analysis of friend responses shows that Twitter users tend to communicate mostly within their cultural regions. Usage of privacy settings showed that privacy perceptions differ across cultures. Finally, we created and evaluated movie recommendation strategies considering user cultural groups, and addressed a cold-start scenario with a new user. We believe that the findings discussed give insights into the sociological and web research, in particular on cultural differences in online communication.

Contents

Chapter 9

Culture-aware Social Recommenders

> "Adapt what is useful, reject what is useless, and add what is specifically your own."
>
> - Bruce Lee, https://www.goodreads.com

This chapter's ideas are based on the publication "Mining Microblogs to Exploit Culture-Awareness in Web Adaptation", co-authored with N.K.Taylor and Y.Jing, and presented at the SICSA PhD Conference 2015 in Glasgow, June 2015.

In Chapters 6 "User Origin Prediction" and 7 "Communication Preferences" we demonstrated a machine-learning approach for inferring tweet country origins based on openly available user profiles. For country-specific adaptation an inferred country location might be exploited in personalisation scenarios. Personal movie preferences might be related to a cultural origin. However, it remains to be seen if movie recommendations with cultural origin information mined from social media would outperform other strategies where this information is not readily available. We propose to exploit inferred tweet locations for movie recommendations based on IMDB movie ratings extracted from user-generated content. Building on previous recommender-system related research works outlined in this chapter, we describe our experimental setup and results achieved, emphasising benefits and drawbacks of our approach.

9.1 Introduction

A widely used movie Recommendation System (RS), such as provided by Netflix[1], predicts user ratings of unseen movies based on previously rated movies. The movie recommendation list is sorted in accord with calculated ratings, showing the highly rated items in the beginning. Predicting movie ratings is a challenging task, in particular when there is limited information on user tastes, characteristics or user feedback on watched movies is not yet available. For improving recommendation outcomes, a RS can employ additional information on user or movie traits. Financial gains when including different user and item-related traits such as user location and music genres into context-aware RS were studied in [5], exploiting matrix factorization models. The findings showed the importance of exploiting user locations for predicting user interests of music concerts, that considering all contextual traits resulted in the highest prediction accuracy and economic value [5].

Despite the potential advantages of including user or movie context information, there are challenges related to particular methods for context-awareness due to different sparsity levels. As discussed by Agarwal with co-authors [8], the number of users and recommended items might differ across contexts, in which larger contexts (with greater number of users/items) can "overpower" predictive performance in small contexts, negatively influencing recommendation outcomes. To deal with "sparse contexts" and challenges of sharing the same factors in different settings, Agarwal with co-authors [8] proposed a recommendation approach considering "local" context-specific factors extending matrix factorisation. In the case of culture-aware recommendation settings, some cultural groups are more likely to be present in the ratings dataset. Thus minority groups would be recommended with items mostly referred by the majority group. This might result in a movie recommendation of top movies without respecting locality or languages.

Other approaches of adapting recommendations to user contexts include pre-filtering and post-filtering for removing ratings unrelated to selected context before and after an employed recommendation technique [185]. However, imposing

[1]https://www.netflix.com

location-specific information filters might be detrimental to the coverage of recommendations and therefore restrict content shown to users. The case of information filtering application for building location-based recommendations becomes more apparent in the following situations. In the movie recommendation setting, restricting user choices based on user locations might considerably limit user choices and result in recommendations which are not so new for the user. While in other recommendation systems such as location-specific news and entertainment events, limiting to specific locations could be desirable for the end user.

It is therefore reasonable to assume that recommending movie and television programs requires knowledge on user location and languages, next to demographic data and history of previous ratings. When the user is new to the system, RS might just exploit user locality traits when available, for instance, from social media websites. For instance, Twitter microblogs can be used to extract user locations out of user-generated content as described by Hecht with co-authors [104]. In previous chapters 6 and 7 we proposed a country predictive models based on user microblogging behaviour patterns [121]; and free-text location field, language and timezone metadata available in open Twitter profiles [52]. Moreover, Twitter streams could provide additional information such as social popularity as exploited in [143] using the Singular Value Decomposition-based (SVD) algorithm with implicit feedback for predicting user ratings. SVD is also used by Rowe [206] considering dynamics of user tastes towards movie genres. Regarding previous ratings history, Dooms with co-authors [60] proposed to collect IMDB movie ratings from Twitter micro-posts, which resulted in follow up of other research works into application of various movie recommendation approaches. Therefore, Twitter microblogs can be used not only as an experimental data source for user opinions on movies, but also to provide an additional data on user traits, extracted or inferred out of microblogs.

However, IMDB ratings posted via Twitter might be biased towards well-voiced and positively rated movies and shows. Will an experimental setup based on the IMDB ratings collected from Twitter be useful for running recommendation experiments despite of the high sparsity and possible rating biases towards most-popular movies? Additionally, predictive country modeling might not lead to performance gains when considering a short-term data collection time-span with no information on user demographics such as age or gender. The effect of including inferred locations might be insignificant when compared to other features, for instance, movie average ratings which might reveal higher rated movies while missing out personal user preferences. We are interested to find out whether our country and cultural region inference improves performance of movie recommendation in absence of users' demographic characteristics as gender and age. We compare culture context-aware RS strategies, build on regression techniques and factorization machines proposed by Rendle [199]. We evaluate RS performance based on movie ratings predicted with and without information on user and movie traits. Our contributions include the following:

- Recommender system experimental setup with IMDB movie ratings collection from Twitter microblogs and collection of user and movie traits from Twitter profiles including user languages and inferred locations; IMDB records including movie genres, runtime and votes number amongst others;
- Analysis of genre preferences for users coming from the top cultural regions and countries as defined in accord the sociological model of cultures proposed by Richard Lewis [148];
- Comparative performance analysis of RS strategies including average-based baselines, culture-aware strategies based on inferred locations, and using the whole feature set with additional user and movie traits;
- Recommendations in performance evaluation of culture-aware RS in light of previous RS studies considering different RS operation modes such as cold-start mode and pre-filtering in accord with inferred user origins.

Next, we briefly outline previous research works pertinent to context-aware RS and exploitation of social media sites. After discussing previous research gaps, we describe and evaluate our recommendation system prototype, concluding with the performance results, further research and industry implications.

9.2 Background

The previous research [180, 262, 225, 80] points out cultural behaviour differences of users online and suggests culture-aware adaptive applications. It remains however unclear whether knowledge on user cultural background could help in improving recommendation performance as compared to country-related recommendations in certain application domains. For instance, some IMDB users might be more interested to different shows and movies which are not really limited to their residence locations, while others might be solely interested in national entertainment options.

Panniello and Gorgoglione [184] investigated applications of contextual pre and post-filtering, user modeling approaches in different shopping situations, whether customers seek presents or shop for personal goods, taking into account the time of the year [184]. The contextual post-filtering idea is to apply a context filter after actually computing the recommendations, and could lead to better results as compared with pre-filtering. It is advised to carefully select post-filtering approaches when simple RS without additional context data outperform pre-filtering [184]. Moreover, context-awareness requires context information to be available.

9.3 Research Questions

There are several issues to be addressed in this chapter. We want to understand whether we could find out location-specific user preferences towards movie genres and if including user context information such as country locations or cultural dimensions could have a positive impact on movie recommendations. For instance, could pre-filtering and factorisation machines be used to exploit the benefits of the additional information on user origins inferred out of Twitter profiles? Could we exploit the social media content in cold-start situations when user is new to the system? Next, in order to understand whether user cultural context could be useful in movie recommendation scenarios, we define the following research questions:

- RQ 4.1: Could we find statistically significant movie genre preferences in relation to user inferred origins?

- RQ 4.2: Could we improve movie recommendation performance when considering user origins or user-related features?

9.4 Experimental Setup

First and foremost, we extract and mine movie ratings and user context information out of openly available Twitter streams. The movie ratings collection is performed similarly to the approach as described by Dooms with co-authors in [60]. Additionally, we store Twitter user profiles data associated with the tweet's author. We use user language, location and timezone for further country predictive user classification model, which allows us to infer tweets' country origins and associated cultural regions.

We also compute "user influence" metric by computing fraction of user followers' from the total number of user contacts. The IMDB related data we collect includes the number of IMDB user votes, IMDB genre (we retain only main genre for our experiments) and the runtime of movies.

Since user locations on Twitter are mostly missed or not accurate [104], we exploit previously introduced cultural dimensions and country predicting model for adding an additional culture-specific user information (please refer to chapter 7). This data is further analysed and pre-processed to create several recommender strategies, including the most recent factorisation machines [199] and gradient boosting regression. Therefore, we address the movie recommendation problem as a machine-learning task of ratings prediction. While regression models help us to factor in additional locality-specific features, we also experiment with context pre-filtering in offline tests.

9.4.1 Methodology

Overall, we perform the following steps:

- for a half of year we collect movie ratings using Twitter search API as explained in [60] while predicting user countries and thus associated cultural dimensions (using the $PLACE$ feature set used for building country predictive models in chapter 7);

- analyse movie ratings (we extract movie ratings shared by Twitter users in the collected tweets) dataset and find out culture and country-specific genre preferences;
- using the movie ratings dataset, compare a simple non-personalised average-based baseline recommendation approach with several recommendation strategies using different feature combinations including user and movie labels, adding user locality and movie-related features such as movie genres for building context-aware recommendations;
- based on rating prediction outcomes considering test performance metrics such as RMSE and NDCG, and cross-validation score of R-squared, find out the best performing strategies in offline recommendation tests.

When addressing RQ4.1, we perform independent t-tests while comparing ratings assigned to movies within the genre and country/culture groups. While answering RQ4.2 we also perform two-tailed t-tests using paired samples, and positively accept the hypothesis that user locality traits inferred out of social media help in improving personalised recommendation approach when at least one of the three aforementioned performance metric values outperforms the average-based baseline recommendation.

9.4.2 Data Collection

With Tweepy Python library, we searched tweets containing a string "I rated IMDB" and stored them into MySQL database when the inferred country was within the list of 15 selected countries (in accord with our country predictive model's performance, see Table H.2). For each of the rated IMDB movies, we also extracted movie-related features such as the primary genre using IMDBpie Python library. It is important to mention that we extracted user-related features while accessing user profiles with Tweepy and further predicted microblog's origin.

Feature	Description	Data Type
VOTES	Number of votes at IMDB	Integer
RUNTIME	Runtime in seconds	Integer
GENRE	IMDB genre, the main genre assigned	Text String
GENRE_AVG	An average movie rating for the respective inferred country and movie genre	Real Number

Table 9.1: Movie features for rating predictive modeling
(three first features retrieved using IMDB API with help of imdb-pie Python Library, the last one is created based on inferred country statistics)

Using information published by IMDB website, we considered several movie features showed in Table 9.1. We assumed that the number of IMDB votes could have an influence on user ratings, and user preferences might also differ for particular movie lengths (runtime). Genres were considered in the aspect of culture-specific user preferences. It is also important to mention that we consider the main genres assigned at IMDB. We observed that Drama, Action, Comedy, Crime and Biography are the top most frequently rated by Twitter users in our dataset.

Table 9.2 presents the user features extracted from Twitter. The WEEKEND feature denotes whether movie rating was posted on weekend, LANGUAGE (user language) and LISTED (the number of Twitter lists) were extracted out of Twitter profile, while COUNTRY was inferred with help of the location-predictive classification model and cultural DIMENSION was assigned for each inferred country using dictionary presented as Table E.1 in appendices. The INFLUENCE is a widely used metric showing fraction of user followers to the total size of user network.

9.4.3 Inferring Countries

For building culture-aware recommendation strategies, and comparing user genre preferences in respect to user cultural origins, we require to know user origins which we mine out of Twitter profiles. We re-use the feature set of the previously created country-predicting classification model developed in chapter 7 “Communication Preferences” (page 122). The classification model was built with location free-text field, timezone and user language extracted from Twitter (comprising $PLACE$ feature set) to create character-based n-gram features, which are represented as a count

Feature	Description	Data Type
WEEKEND	Extracted from the posting date	Boolean, equals to True (1) when posted on weekend. Otherwise False (0)
LANGUAGE	User language defined in Twitter profile	Text String
COUNTRY	User country predicted using LOCALITY feature	Text String
DIMENSION	User cultural dimension associated with the inferred country using dictionary	Text String
INFLUENCE	Ratio of followers to the total network size	In percent, Integer
LISTED	Number of lists into which this user profile is included	Integer

Table 9.2: User features for rating prediction models
(INFLUENCE and LISTED features are derived from user profile using Tweepy Python Library)

matrix and further converted into its normalised Term Frequency-Inverse Document Frequency statistics. Instead of using Multinomial Naïve Bayes classification, we exploited character-based Linear Support-Vector Classification[2], which enabled lower test error than word-based and smaller number of features as compared to model unifying the both feature sets (Figure H.6 on page 270 in appendices).

9.4.4 Recommendation Experiments

To address the second research question, we compare several recommendation strategies' outcomes, while considering user preferred language defined in Twitter profile, automatically inferred tweet origin country and associated cultural group. We thus create a hybrid recommendation engine taking into account user cultural origins, social networking traits and movie-related traits.

Evaluation process and experimental goals In order to get insights into culture-aware recommender performance, we perform series of experiments with offline data extracted from Twitter streams. We are interested to learn whether information on tweet origins would be useful for providing better recommendations when

[2]using `http://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html`

compared with simple average-based recommendation baseline and non-context enhanced regression models. While comparing different recommendation approaches and context data inclusion strategies, we evaluate ratings predictions aiming at a smaller prediction error. We also consider ranking accuracy for top 10 recommendations list and goodness of regression fit in cross-validation tests as described further. It is important to mention that we separate Cross-Validation, test and training sets in all our experiments. We use separate training and testing sets also when searching for regression model parameters.

Performance Metrics To evaluate an accuracy of recommendation systems, several information retrieval, decision-based and ranking based metrics are employed in respect of recommendation system objectives [107]. Decision-based metrics such as Precision and Recall arguably are challenging to apply to the RS evaluation tasks due to the non-binary ratings and their natural subjectivity for defining real user interests and thus assessing the recommendation relevance in practice. Receiver Operating Characteristics (ROC) curve is alternative to decision-based metrics above and is suitable when it is required to find all good items in binary recommendation/prediction tasks. However, ROC curve usage requires a considerable test set for each user [107]. Therefore ROC curve is not applicable to our (very) sparse ratings dataset.

In this thesis, we employ Root Mean Square Error (RMSE), which is widely used in practice and theory for evaluating the accuracy of predictive or recommendation systems, for instance, it was used for comparing the movie recommendation performance in Netflix grand prize competition [207]. The winning solution (see [136]) achieved a 10% improvement over the defined baseline [3]. RMSE measures the differences between predicted and observed ratings and requires to be minimised when possible [207]:

$$RMSE = \sqrt{\frac{\sum (\hat{y} - y)^2}{n}}, \tag{9.1}$$

[3] `https://www.netflixprize.com/community/topic_1537.html`

where $\hat{y}$ is the predicted rating, y is actual rating, n is total number of ratings.

Ideally, the list of the movie recommendations is sorted from the "most interesting" top movie to the "least interesting" for the user movie. This way users can focus on the highly ranked movies located on the top of the recommendations list. To assess the quality of the ranking, Discounted cumulative gain (DCG) is widely used in web applications including search engines [266]. As explained in [266], DCG value can be normalised by dividing it by the Ideal Discounted Cumulative Gain (IDCG) denoting perfect ordering of the recommendations list. For our calculations, we use the formula for NDCG at position N (assuming that users are interested in top-N movie items) as follows [31, 266] :

$$NDCG@N = \frac{DCG@N}{IDCG@N}, \tag{9.2}$$

$$DCG@N = \sum_{i=1}^{N} \frac{2^{r(i)} - 1}{\log_2(1+i)}, \tag{9.3}$$

in which, applicably to our movie recommendation task, we consider N=10 in all our tests (however, in practice users could be provided with more recommendations), i is i-th element in movie list (sorted by decreasing rating score) and $r(i)$ is its relevance (or rating). We can achieve the best ranking performance value when NDCG equals one.

Moreover, since we test and evaluate several regression models, which are required to be provided with their own parameter settings, it was important to consider the metric we use for cross-validating different parameter combinations. We want to understand how much of the variation in user ratings can be explained by our model with the tested parameters and context inclusion strategies. For this, we compute the coefficient of determination as follows[4]:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}, \tag{9.4}$$

where $\hat{y}_i$ is the predicted value of the i-th movie rating and y_i is the corresponding

[4] `http://scikit-learn.org/stable/modules/model_evaluation.html`

true value, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the average movie rating.

Overall, we exploit R^2 for models' comparison and parameter tuning in CV tests. While performing tests on separate testing sets, we aim to minimise the RMSE value while considering higher values of $NDCG@10$ and R^2 while selecting recommendation strategies which outperform non-personalised baseline such as based on rating averages. We employ $NDCG@10$ due to the assumption that users might prefer to see the higher rated items in the top-10 list of recommendations.

Parametrisation and Evaluation steps. Having about 19 weeks of data collection and several regression models, it was necessary to tune the model parameters. For that, from the whole movie ratings table "ALL" we randomly select a small sample of 7K Twitter user ratings (the data table "SAMPLE" is summarised in Table J.1). To evaluate their performance, we established a set of baseline strategies based on average ratings as follows:

- MU: the simplest rating predictor based on the overall average movie rating;
- $ITEM$: the rating predictor based on the average movie rating of the selected movie;
- $USER$: the rating predictor based on the average movie rating of the selected user;
- $OFFSET$: the rating predictor based on movie rating offsets for particular users and movies with their averages, this strategy is widely used as a baseline as described in [63];

For each rating we retain user and movie labels, also user and item-related traits discussed above. These additional features are excluded from the average-based baseline strategies, based on item and user labels together with their respective average ratings in case of $OFFSET$ strategy. The aim is to understand if the context traits help to achieve a better performance (smaller RMSE or higher NDCG@10/R^2 values) in user modeling strategies (when we enhance recommendation models with these traits) and in cold-start situations when there are little or no user ratings available.

While deciding on the initial parameters ranges for our regression-based models, we performed random search and parameters tuning. The models selected were evaluated on the “TIMELINE” data , which does not overlap with the “SAMPLE” data used for parametrisation experiments and tests. In the final test on “TIMELINE”, we train our tuned models while adding user ratings week by week. The last week is used for testing purposes. This test helps us to analyse how recommendation models’ performance changes when adding more training instances.

Recommendation Strategies Tested. Having data on movie ratings provided by Twitter users and related contextual information described in section 9.2 above, we evaluate the following models with several context attribute combinations defined in Table 9.3.

Strategies	Included user and item features
IDs	herein we do not exploit any additional user or movie attributes besides user ID and movie ID labels
BASE	IDs with added user and movie average ratings
LOCALITY	is built on BASE model and includes inferred country and related cultural dimension (Lewis Model of Cultures[149]), language defined in the user profile
COUNTRY	is built on BASE model and includes inferred country
DIMENSION	is built on BASE model and includes cultural dimension associated with the inferred country
LANGUAGE	is built on BASE model and includes user language defined in the Twitter profile
ALL	includes all user and movie-related features including “DIMENSION”, “COUNTRY”, “LANGUAGE”, “LISTED”, “WEEKEND”, “INFLUENCE”, “VOTES”, “GENRE”, “RUNTIME” and “GENRE_AVG” (an average movie rating for the respective genre and inferred country)

Table 9.3: Context inclusion strategies

Parameter	Description	Tested Values
FACTORS : Random search of parameters in 15 3-times CV tests		
num_iter	number of iterations	[1..100]
num_factors	number of hidden factors (feature interactions)	[1..50]
learning_rate	initial learning rate	0.0001, 0.001, 0.01
BOOSTER: Using boosting iterations, we estimate test deviance loss function		
max_depth	maximum depth of the decision trees (having shallow trees)	3, 5, 7
learning_rate	the minimum number of instances at leaf nodes	0.01, 0.015, 0.025
n_estimators	number of of boosting stages, larger number is generally leads to better performance	min(number of features, 600)
subsample	the fraction of samples used for fitting shallow trees	0.8, 0.9, 1

Table 9.4: Value ranges for model parameters' tuning

For creating context-aware and un-contextual RS models, we consider the following regression techniques (in our pilot tests, we also tested Random Forest and LASSO further disregarded as justified in Appendix J.1):

- *FACTORS* : state-of-the-art Factorisation machines proposed by Rendle [199];
- *BOOSTER*: Gradient Boosting Regressor with least squares loss function, provided by scikit-learn.org team.

Table 9.4 shows the tested initial values and ranges of model parameters. The best-performing parameter settings were exploited for final models creation and testing. The final tests are performed using 5-times CV on the "SAMPLE" and "TIMELINE" data tables. In "timeline" tests, we perform tests while adding user ratings on a weekly basis, resulting in the overall 19 tests which we compare for all recommendation strategies, using paired t-tests.

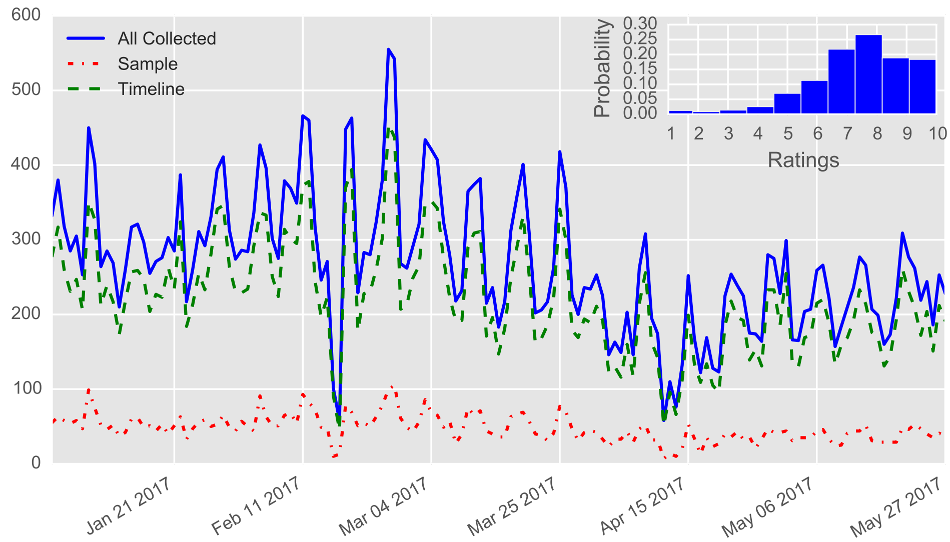


Figure 9.1: Data collection in time
(Collecting tweets with movie ratings using Twitter search interface [242] resulted in 3 data tables with movie ratings as a part of dataset "Recommender" listed in Appendix D: 1. All collected: all tweets, 2. Sample: used for models' tuning, 3. Timeline: for running timeline tests using tuned models. The ratings distribution is skewed towards higher ratings with median of 8.)

9.5 Findings

9.5.1 Dataset

Using Twitter search API, we collected a set of about 39,596 tweets with the search phrase "I rated IMDB movie" from 1st of January till 27th of May 2017. Out of this data table "ALL", we randomly selected 7K of tweets comprising "SAMPLE" and the rest included into the "TIMELINE" table as shown in Table 9.5. The average rating in "ALL" data table was 7.6 ± 1.8 [5]. After running some trial data collection tests, we experienced a steady data collection trend whereas the majority of days had at least 22 hours of movie rating collection with 221 ± 75 ratings per day in average as shown in Figure 9.1.

We can observe that very few movies have a high number of ratings, while very few movies are very often rated, similar "Long-tail" distribution is seen for users in the all three cultural groups in Figure 9.2.

[5] We see here the standard deviation after "±" sign

Table	Ratings	Users	Movies	Countries Inferred	User Languages
SAMPLE	7000	3001	2411	15	17
TIMELINE	32596	6349	6711	15	20
ALL	39596	6891	7446	15	20

Table 9.5: Rating tables and unique names counts
(ALL: all collected tweets, SAMPLE: used for models' tuning, TIMELINE: for evaluation tests)

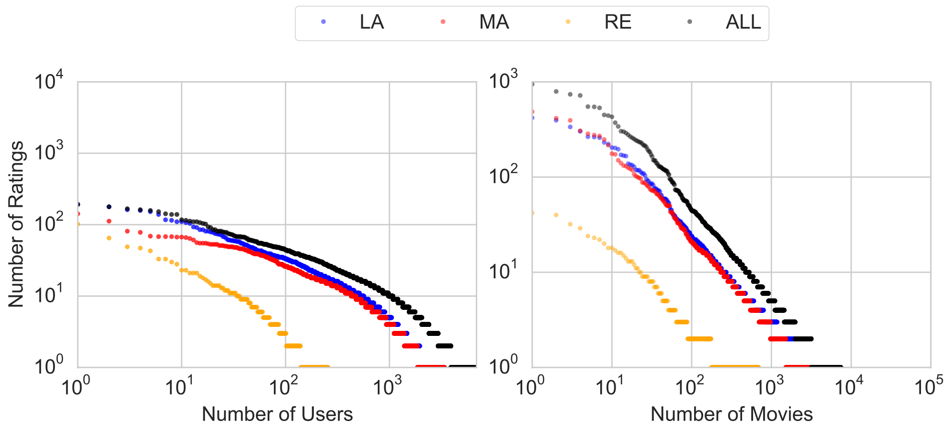


Figure 9.2: Number of ratings log distributions shows ratings of users and movie ratings (ALL: followed users are grouped into LA: Linear-active, MA: Multi-active, RE: reactive in accord with the Lewis Model's cultural categories [148])

9.5.2 Tweet Origin Prediction

We used our country-predictive classification model using PLACE feature set based on free-text location field, timezone and user language defined in the Twitter profile (Table J.2 shows country inference examples). All tweets were assigned with their respective inferred 15 countries (we omitted other tweets while collecting dataset) including following shown in Table 9.6

In the "ALL" ratings table we observed only 42 messages with country location information out of 39596 Twitter messages in total. Since our goal was to find out a match of defined country locations with the results of the country predictive model, we extended the geo-tagged tweets set with all tweets we collected during the trial data collection tests. Overall, we counted 432 tweets published by 56 users used for creating a classification report showing an overall accuracy of 75% as seen in Table 9.5.2. However, when looking at particular country classes, such as Japan or Thailand, we observe unsatisfactory classification performance. This happens due to unbalanced countries, as GB and US "out-powers" other country groups. For

Dimension	Country (ISO Code [a])	Users	Ratings	μ_{rating}	σ_{rating}
LA	United States (US)	2547	14249	7.60	1.89
LA	Great Britain (GB)	783	6213	7.46	1.62
LA	Sweden (SE)	74	455	7.07	1.96
MA	Saudi Arabia (SA)	2243	11228	7.99	1.91
MA	Turkey (TR)	430	2813	7.18	1.65
MA	Russian Federation (RU)	407	1060	7.10	2.02
MA	Italy (IT)	96	680	7.15	1.53
MA	Brazil (BR)	88	539	7.58	1.81
MA	Spain(ES)	84	495	7.61	1.80
MA	Venezuela (VE)	22	233	7.36	1.66
MA	Greece (GR)	22	149	6.77	1.98
RE	Indonesia (ID)	62	638	7.39	1.74
RE	India (IN)	95	457	7.84	1.79
RE	Japan (JP)	56	211	7.45	1.74
RE	Thailand (TH)	45	176	8.16	1.33

Table 9.6: Inferred countries and their average (μ) movie ratings with standard deviations (σ)

[a]ISO codes available at `http://www.nationsonline.org/oneworld/country_code_list.htm`

instance, almost 40% of Twitter users predicted as tweeting from the USA, and from other locations including Great Britain (3.12%), Greece (6.25%) , Japan (3.12%), Turkey (3.12%), India (3.12%), Saudi Arabia (15.62%) and Thailand (6.25%).

This information was used in an additional human assessment of the country predictive model by two raters including the thesis author. Next, we report the rater agreement percentage and the agreement reliability. In Methodology chapter 5.3.2 we provided an overview and discussed usage of several Inter-Rater Reliability (IRR) coefficients. With help of Python libraries we computed Krippendorff's α [91], Cohen Kappa [213] and Fleiss' Kappa [192] with the results in appendices, chapter H.2.

During the human assessment test, raters were provided with languages defined in Twitter user profiles, their free-text location descriptions, time zones when enabled in the user device. To make the assessment easier for the raters, we included also links to all 56 Twitter profiles, tweet content with the movie rating, "about me" information retrieved from the Twitter [6]. The country locations, both inferred and

[6]For privacy reasons, we removed this data from the Table H.3 in appendices

Country ISO Code (Country name)	Precision	Recall	F1-score	Support
ES (Spain)	1.00	1.00	1.00	2
GB (Great Britain)	1.00	0.92	0.96	13
GR (Greece)	0.00	0.00	0.00	2
ID (Indonesia)	1.00	1.00	1.00	3
IN (India)	0.00	0.00	0.00	1
JP (Japan)	0.00	0.00	0.00	1
RU (Russian Federation)	1.00	1.00	1.00	4
SA (Saudi Arabia)	1.00	0.17	0.29	6
TH (Thailand)	0.00	0.00	0.00	2
TR (Turkey)	1.00	0.50	0.67	2
US (United States)	0.59	0.95	0.73	20
average / total	0.75	0.75	0.70	56

Table 9.7: Country prediction classification report comparing automatic country detection with the data provided by the user device
(using classification report provided in Sklearn Python library with metrics.classification_report)

provided by user devices, were not initially available for the raters' review. First of all, raters independently assigned country codes for 56 of users (1 user account was not available, possibly suspended by Twitter). Secondly, raters reviewed the ratings together and in 52 cases out of 56 users agreed on the country origins of 56 users (93% of cases). Table H.4 (a) shows Krippendorff's $\alpha = 0.94$ (corresponds to very good reliability outperforming the benchmark level of $\alpha = 0.80$ advised by Krippendorff in [140]), while (b) shows Cohen and Fleiss Kappas of equal values $\kappa = 0.92$ (almost perfect results in accordance with the benchmark scale by Landis and Koch [145]).

Table H.4 (a) shows that Twitter country locations provided by the user device agreed with the country-predictive model in 42 of cases (75%), which was slightly higher than the match of human raters and the predictive model (70% and 68%). Human raters agreed with the country labels provided by Twitter in 84% and 79% of cases. This demonstrates that to achieve the model match with both raters is quite challenging as compared with the output of a Twitter device, proving a reasonable source of labelling information for automatic creation of supervised country-predictive models.

It is important to reiterate that we computed all coefficients to monitor their possible disagreement as discussed in the chapter 5 “Methodology” referring to the critique by Zhao [281]. We observed that in most of the cases the conclusions based on these three coefficients agreed, except of the case of Krippendorff’s α reaching less than moderate results for IRR between Classifier and Human Rater1 $\alpha = 0.60$, Classifier and Human Rater2 $\alpha = 0.59$ and Classifier and Twitter $\alpha = 0.66$ (below Krippendorff’s α threshold of 0.67 for ”tentative conclusions”). Table Table H.4 (b) presents Fleiss’ and Cohen Kappa coefficients of moderate and good values (three values for both coefficients are in range from 0.58 to 0.65) in accord with the benchmark scale by Landis and Koch in [145]. Table H.4 (c) with IRR results for more than 2 annotators showed considerable results for Fleiss’ Kappa and Krippendorff’s α in range from 0.7 to 0.82 and from 0.72 to 0.84 respectively. The highest values of Fleiss’ Kappa of 0.82 and Krippendorff’s α of 0.84 were for Human raters and Twitter country annotations. All four annotators, human and automatic (Twitter and country predictive model) have reached substantial agreement levels in accordance with the benchmark scales aforementioned, with Fleiss’ Kappa of 0.72 and Krippendorff’s α of 0.74 as shown in the last row in Table H.4 (c).

9.5.3 Movie Genre Preferences

The previous research [5] findings using user survey reveal that persons coming from different cultural groups have their own genre preferences. Assuming that movie genre preferences might also differ across inferred location groups, we performed independent t-tests for comparing the movie rating means for the inferred locality groups. In total, we had 25 movie genres assigned to all rated movies from our dataset. Next, we compared user ratings for the top 5 most referred genres in the “ALL” data table (Table D.6), including Drama (11135 ratings), Action (8981 ratings), Comedy (5490 ratings), Crime (3543 ratings) and Biography (2635 ratings). Users were broken into groups in accord to with their inferred country, user language defined in user profile and cultural dimension associated with the inferred country. The ratings of users from the defined user groups were compared to the rest of users. Based on t-test results, significant rating differences (with $p < 0.05$ and $p < 0.01$

levels) were found in respect of user origins and genres of rated movies presented in Appendix I.

It is interesting to note that a majority of persons with defined Arabic in their user profiles (language-based user group) and persons with inferred country of Saudi Arabia tend to rate Drama movies in average higher as compared with other respective locations (test #53 in Appendix I: for Arabic language in the user profile $\mu_1 = 8.11$ and $\sigma_1 = 1.94$ versus other $\mu_2 = 7.66$ and $\sigma_2 = 1.76$; test #38: for inferred country as Saudi Arabia $\mu_1 = 8.07$ and $\sigma_1 = 1.93$ versus other $\mu_2 = 7.64$ and $\sigma_2 = 1.75$) with $t(4649.44$[7]$) = 10.73, p - value < 0.01$ and $t(6432.82) = 10.95, p - value < 0.01$ respectively, while persons inferred as coming from Turkey or having Turkish language defined in user profile tend to rate Drama movies significantly lower (test #55: for Turkish language in the user profile $\mu_1 = 7.05$ and $\sigma_1 = 1.64$ versus other $\mu_2 = 7.83$ and $\sigma_2 = 1.82$, test #39: for inferred country as Turkey $\mu_1 = 7.22$ and $\sigma_1 = 1.64$ versus other $\mu_2 = 7.84$ and $\sigma_2 = 1.83$) with $t(10,291) = -10.34, p - value < 0.01$ and $t(10,291) = -9.88, p - value < 0.01$ respectively. It is interesting to learn that even though that localities belong to the same cultural dimension of Multi-Active persons, their preferences for the Drama genre differ considerably. We further explore whether cultural origins of Twitter users (which are also users of IMDB database) have some effects in several movie recommendation scenarios.

9.5.4 RS Performance

Parametrization Outcomes. As described above, we performed 3-fold CV while finding model hyper-parameters in random search for $FACTORS$ model, while $BOOSTER$ model parameters were found using "Out-of-bag" estimates. Table 9.8 shows the found parameters for the selected context inclusion strategies. We further exploit these parameters in the final tests on "TIMELINE" data and user offline tests.

[7] Herein for tests #53 and #38 we report the degrees of freedom with Welch-Satterthwaite correction used with Welch's t-test, not assuming equal population variance (based on the Levene's test output with $p - value < 0.05$).

Context Strategy (parameters)	$\mu R^2 \pm \sigma R^2$	RMSE	NDCG@10
BOOSTER: gradient boosting regression			
COUNTRY (n estimators=600, subsample=1, learning rate=0.01, max depth=5)	0.75±0.03	1.79	0.59
ALL (n estimators=600, subsample=1, learning rate=0.015, max depth=3)	0.74±0.03	1.77	0.59
LANGUAGE (n estimators=600, subsample=1, learning rate=0.015, max depth=5)	0.74±0.03	1.79	0.56
DIMENSION (n estimators=600, subsample=1, learning rate=0.015, max depth=5)	0.74±0.03	1.79	0.62
BASE (n estimators=600, subsample=0.8, learning rate=0.025, max depth=3)	0.73±0.03	1.77	0.60
LOCALITY (n estimators=600, subsample=0.9, learning rate=0.025, max depth=3)	0.73±0.03	1.77	0.59
IDs (n estimators=600, subsample=0.8, learning rate=0.025, max depth=7)	0.10±0.02	1.71	0.73
FACTORS: factorisation machines			
COUNTRY (num factors=42, learning rate=0.0001, num iter=95)	0.72±0.03	4.64	0.32
DIMENSION(num factors=16, learning rate=0.0001, num iter=67)	0.72±0.02	3.78	0.54
ALL (num factors=17, learning rate=0.01, num iter=1)	0.71±0.03	6.90	0.44
LANGUAGE (num factors=23, learning rate=0.0001, num iter=71)	0.71±0.02	4.17	0.73
BASE (num factors=5, learning rate=0.001, num iter=16)	0.65±0.15	4.95	0.29
LOCALITY (num factors=42, learning rate=0.001, num iter=66)	0.61±0.18	5.06	0.74
IDs (num factors=42, learning rate=0.001, num iter=93)	0.17±0.02	1.86	0.51

Table 9.8: Tuned model parameters and performance using movie ratings from "SAMPLE" data table (first column shows the tuned model parameters, second column shows R^2 in 5-times CV, two last columns show RMSE and NDCG@10 using separate test set)

We observed that parameters search improved the overall R^2 performance for the both models within the all context feature sets as shown in Figure 9.3. The most importantly, BASE model's R^2 was comparable with the all context-aware strategies as result of the regression models' tuning.

Cross-Validation using "SAMPLE" set. Using the parameters selected for each of the context strategies, we performed 5 times cross-validation tests with the "SAMPLE" data. Figure 9.4 shows that the inclusion of user and movie averages

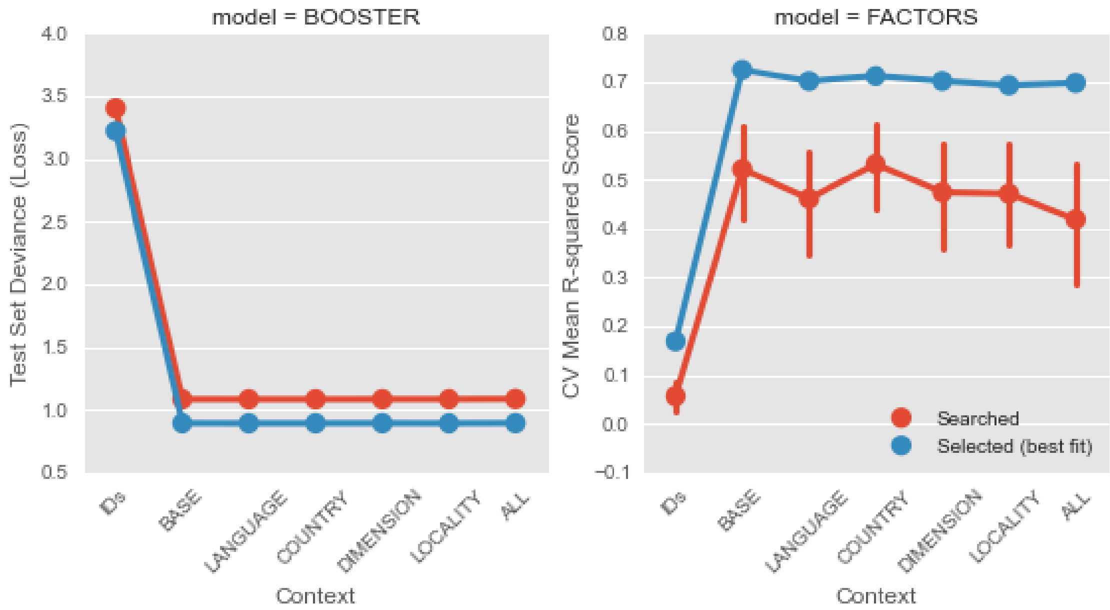


Figure 9.3: Parametrisation results performed on a small "SAMPLE" movie ratings set, mean performance with 95% confidence intervals
($BOOSTER$: Gradient Boosting Regression Model was tuned by minimising the regression error on the test set, $FACTORS$: Factorisation Machines was parametrised using random parameters search with 3 times Cross-Validation in 15 iterations)

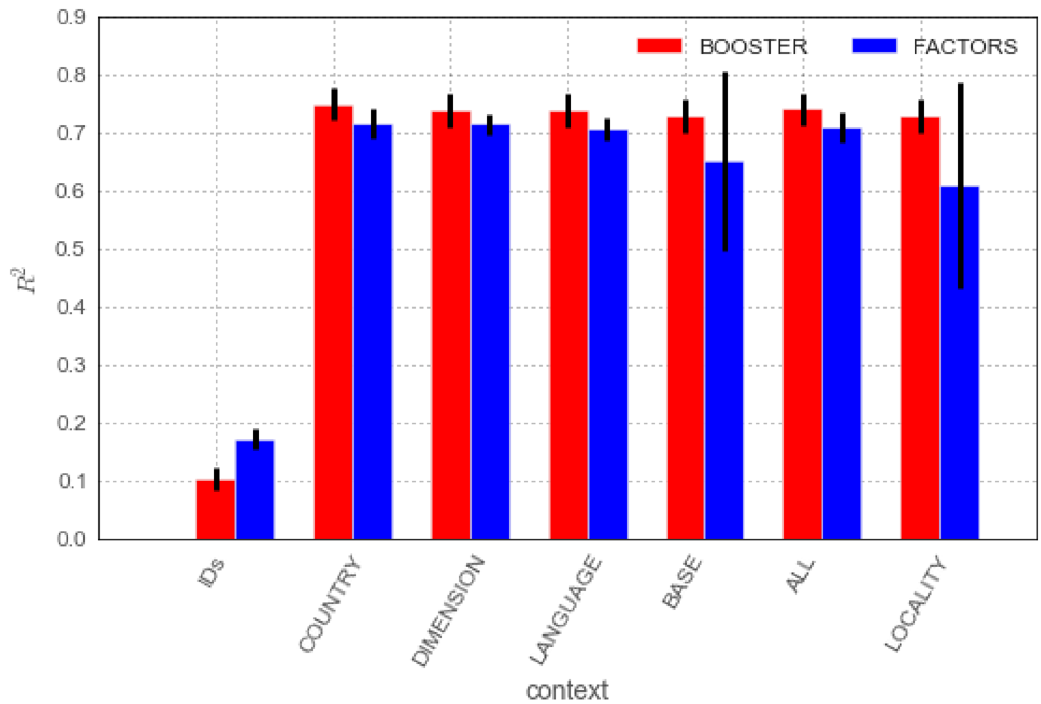


Figure 9.4: Average R^2 values with st.dev. errors in 5-times cross-Validation tests on "SAMPLE" data ($BOOSTER$: gradient boosting regression model was tuned by minimising the regression error on test set, $FACTORS$: factorisation machines was parametrised using random parameters search with 3 times cross-validation in 15 iterations)

helped to improve the coefficient of determination R^2 and thus fraction of variance explained increased when compared with the "IDs"-based models. The means for R^2 in "BASE" models do not differ significantly from the models using inferred from Twitter location features. The standard deviation is considerably larger for $FACTORS$ model when employing "BASE" and "LOCALITY" strategies as compared to other model settings. Even though we might disregard all locality-based context strategies, it is important to mention that the tests were performed on a relatively small sample of 7K ratings. We further analyse how these models' performance metrics change when adding more ratings on a new "TIMELINE" data. Since we get the lowest R^2 mean values for "IDs" context strategy, we disregard it in further timeline tests.

Baseline Recommender Selection. To assess performance of culture-aware recommendation strategies, we considered several non-personalized baselines based on average user and movie ratings. We tested several average-based baseline strategies including $OFFSET$, USER, ITEM, and MU described above and depicted in Figure J.8 in appendices. The $OFFSET$ was of comparable NDCG performance with USER recommendation strategy (p-value = 0.56, USER $\mu = 0.81 \pm 0.11$, OFFSET $\mu = 0.83 \pm 0.09$). In all other cases, the $OFFSET$ baseline significantly outperformed USER, ITEM and MU for all tested performance metrics (p-value < 0.05), therefore, we consider the $OFFSET$ for further comparison of the context-aware recommendation strategies.

Context-aware Recommendations. Figures J.9 ($RMSE$), J.10 ($NDCG$) and J.11 (R^2) in the appendices show performance of the context-aware strategies ($BOOSTER$ and $FACTORS$) using "TIMELINE" data. The $RMSE$ performance is comparable (not significantly different with p-value > 0.05) for $OFFSET$ and all $FACTORS$-based strategies except "ALL" (when using all user and item-related features). The $RMSE$ performance for $FACTORS$:ALL strategy is significantly higher than $RMSE$ for $OFFSET$ ($p - value < 0.05$, $FACTORS$ $\mu = 1.74 \pm 0.06$, $OFFSET$ $\mu = 1.69 \pm 0.07$). In opposite, the $RMSE$ performance of all $BOOSTER$ context-aware strategies outperforms $OFFSET$ with statistical significance of $p - value < 0.01$.

Figure J.10 shows that the $NDCG$ performance is comparable for all strategies in $FACTORS$ and $OFFSET$ models. In opposite, all $BOOSTER$ context-aware strategies outperform $OFFSET$ in the $NDCG$ performance with significance ($p - value < 0.01$). We can see trend lines for the $OFFSET$ model do not show a positive increase for $NDCG$ values as we add more ratings week-by-week. The "BASE" average-based strategy shows slightly negative trend for ranking performance reflected in the NDCG metrics values. However, context-aware recommendation strategies for $FACTORS$ model, such as COUNTRY, ALL, LOCALITY, DIMENSION, demonstrate that $NDCG$ performance can be further improved in time.

As seen in Figure J.11, however, all R^2 performance lines show negative trends for both regression models. All context-aware strategies for $BOOSTER$ model outperform $OFFSET$ in R^2 with statistical significance ($p - value < 0.01$). Similarly, we observe that $FACTORS$ model significantly ($p - value < 0.01$) outperforms $OFFSET$ model in almost all context-aware strategies except "LOCALITY" ($p - value \approx 0.1$).

Performance Improvement over the Baseline. We observed that inclusion of locality traits inferred out of Twitter microblogs can help in achieving better performance values. Compared with the $OFFSET$ average-based predictive model, inclusion of additional data gathered from Web and exploitation of Gradient Boosting ($BOOSTER$) regression model led to an improved R^2 (with statistical significance $p - value < 0.01$) by 41.96% from .45±.05 to .65±.06 using $BOOSTER$:COUNTRY (see Table J.5 case #39), followed by $BOOSTER$:ALL (case #35, improvement of 41.35%), $BOOSTER$:LANGUAGE (case #38, improvement of 40.56%), $BOOSTER$:DIMENSION (case #41, improvement of 40.54%), $BOOSTER$:LOCALITY (case #40, improvement of 39.22%). Similarly, we observed R^2 improvement (from 30.36% to 32.99% with $p - value < .01$) for context-aware $FACTORS$ strategies including COUNTRY, LANGUAGE, DIMENSION and ALL. The $FACTORS$:LOCALITY showed the R^2 improvement of 12.03% (case #33, with $p - value < 0.05$). The BASE strategy in $BOOSTER$ and $FACTORS$ models showed the improved R^2 by 39.37% (case #37) and 27.33% (case #30) respectively ($p - value < .01$). The IDs-based strategies could not compete with the baseline and showed the considerably worse R^2 performance (with $p - value < .01$ for both cases #29 and #36).

The NDCG's statistically significant increase (with $p - value < .01$) over the $OFFSET$ baseline was observed for $BOOSTER$ strategies including LOCALITY (by 17.94%, case #26), DIMENSION (by 17.07%, case #27), COUNTRY (by 17.10%, case #25), ALL (by 16.55%, case #21), LANGUAGE (by 14.75%, case #24). In opposite, $FACTORS$ did not benefit from the added variables showing no significant differences (all $p - value \geq 0.05$), and even a statistically significant however small decrease in $NDCG$ when we added all possible variable (decreased by 8.69% in case #14, $p - value < 0.05$).

The $BOOSTER$ model achieved a better (decreased by from 12.78% for ALL to 11.75% for LANGUAGE) $RMSE$ performance for all context inclusion strategies, and also for BASE strategy (all with $p - value < .01$). The IDs-based $FACTORS$ strategy showed a decrease in RMSE by 1.73% (case #1), while $BOOSTER$:IDs showed an increase in RMSE by 4.15% (case #8). All the rest of $FACTORS$ strategies showed quite small or statistically insignificant RMSE changes.

It is important to mention that all context-aware strategies included user and movie average ratings as a standard practice. Therefore it was reasonable to compare the BASE feature set with the context-aware strategies. We added locality traits into the BASE feature set using Gradient Boosting regression model ($BOOSTER$). Table J.6 shows that BASE baseline is very competitive for most of our strategies across all performance metrics. For instance, RMSE of $BOOSTER$:BASE is smaller than $BOOSTER$:IDs (case #1) and $BOOSTER$:ALL (case #8) by 19.19% and 17.19% respectively (with $p - value < .01$). However, when can also achieve an improved NDCG ($p - value < 0.05$) when adding all parameters (by 2% in case #14) or using LOCALITY strategy (by 3.21% in case #18). Therefore the inferred locality helped to improve the NDCG performance even with the stronger BASE baseline and using data extracted from Twitter accounts.

Ranking the recommendation stratagies. Further, we ranked [8] the aforementioned recommendations strategies. Table J.3 in appendices summarises the ranks (higher ranks indicate that the related performance metric values are better compared to the other strategies, while lower ranks indicate that the worst performance outcomes). The best context-aware RS strategies include *BOOSTER*:COUNTRY, *BOOSTER*:ALL and *BOOSTER*: LANGUAGE in R^2 values. In test NDCG performance the best ranked strategies include *BOOSTER*:LOCALITY, *BOOSTER*:COUNTRY and *BOOSTER*:DIMENSION. In test RMSE values we observed the best performance ranks for *BOOSTER*:ALL, *BOOSTER*:BASE and *BOOSTER*: LOCALITY. Overall, the mean ranks amongst the all parameters showed the best values for *BOOSTER*:ALL, *BOOSTER*: COUNTRY and *BOOSTER*:LOCALITY as seen in critical differences diagram for the average ranks in Figure 9.5. However, would these strategies be beneficial for individual users? Having a large dataset we might select a users' sample and perform the recommendation "simulation" tests while breaking collected users' movie ratings into test and train sets in the next section.

Offline Tests with Selected Users. We are further interested to know whether the inferred locality might help also in individual user tests. For this, we selected a sample of users having at least 3 ratings in our "TIMELINE" data. With the aim to test pre-filtering approach using cultural dimension group, and also considering the aforementioned context strategies and models (*FACTORS* , *BOOSTER*) in cases of cold-start problem when having no previous user ratings available while training the recommendation model. We created a sample of 270 users, whose average ratings and number of ratings distribution is depicted in a scatter plot in Figure 9.6.

[8] `https://docs.scipy.org/doc/scipy-0.16.0/reference/generated/scipy.stats.rankdata.html`

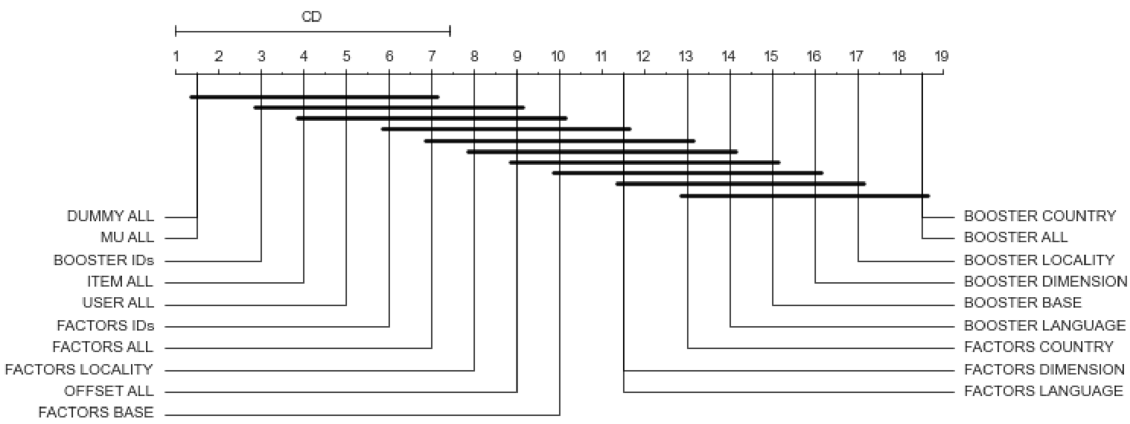


Figure 9.5: Critical difference for ranking of the recommendation strategies (plotted with the help of "stats.rankdata" provided by Python library Scipy using average method for dealing with tied ranks, and the "evaluation.scoring.compute_CD" for computing and plotting the critical difference for Nemenyi test with $\alpha = 0.05$ provided in the Python library Orange)

Figure 9.6: Selected sample of users (270 users depicted in colored circles have rated at least three movies) for further movie recommendation tests

Pre-filtering and locality usage. Table J.4 in appendices shows results of paired two-tailed t-tests using selected users' set. Herein we report statistically significant results with $\alpha < 0.01$ while comparing the tested recommendation strategies with the $OFFSET$ baseline based on average movie ratings. Using the Shapiro-Wilk statistical test for normality we accepted the assumption of normality for pared performance metric differences in 10 out of 90 test cases with $\alpha < 0.05$, we ignored the normality assumption violation having hundreds of observations (from 164 to 270 in all tests) as discussed in [16].

We observed R^2 performance improvement from .43 to 0.48, which accounts for at least 10% increase when using $BOOSTER$:LOCALITY (based on 204 user comparisons, case #41) and by 11% when using dimension-based pre-filtering (164 user comparisons, case #72) thus using only dataset respective to the inferred user cultural group. We also emulated cold-start situation by using all individual user ratings in the test sets (the train sets were empty and using only data of other users). By applying user dimension pre-filtering in cold-start, we observed almost 9% improvement in R^2 values, from .49 to 0.53±.06 with $p{-}value < 0.01$ significance (case #72), and also $RMSE$ decrease by 4% (case #42) from 1.09 to 1.04±.52 with $p - value < 0.05$.

Summary of tests. Table 9.9 summarises the most significant findings from the offline tests with selected users sample test and timeline tests, wherein data was added on a weekly basis. The gradient booster ($BOOSTER$) strategies helped to reach the best performance especially when using ALL, BASE and LOCALITY feature sets. In cold-start, pre-filtering helped to improve RMSE and R^2 performance. Overall, the worst performance was observed for IDs-based recommendation strategies.

Cases	Impact	Findings
		Selected user tests in Table J.4 (pre-filtering approach)
51	-	The RMSE increase by 33% (with significance of $p - value_51 < 0.01$) when using the BOOSTER:IDs strategy with pre-filtering (using cultural dimension) compared with the OFFSET.
9, 2	+	BOOSTER and FACTORS models using BASE feature set showed a decrease by 7.62% and 8.45% respectively ($p{-}value_9 < 0.05$ and $p{-}value_2 < 0.01$) in RMSE performance as compared with the OFFSET.

Table 9.9 : continued on the next page

Cases	Impact	Findings
57, 68, 61, 62, 63, 60, 71, 67	+	When using pre-filtering on inferred user dimension, the NDCG metric was comparable to OFFSET or slightly improved in few cases by almost 2%, from 1.88% to 1.51% (with $p - value < 0.05$).
72	+	R^2 improved by 11.04% when using $OFFSET$ pre-filtering using inferred cultural dimension (with $p - value < 0.01$).
41, 37	+	When using BOOSTER:LOCALITY, R^2 improved by 10.19% as compared to the $OFFSET$ ($p - value < 0.01$), BOOSTER:LOCALITY followed by BOOSTER:BASE with R^2 improvement by 9.83% (for both cases, $p - value < 0.01$).
29, 74, 36, 81	-	IDs-based approaches were poorly performing in R^2 (from 42.91 to 63.10% decrease) for BOOSTER and FACTORS models as compared with OFFSET in usual recommendation and pre-filtering strategies (with $p-value < 0.01$).
Selected user tests in Table J.4 (cold-start cases)		
51, 44, 8, 1	-	IDs-based showed the poor performance in the $RMSE$ values in cold-start test cases, with $RMSE$ increase from 41.71% to 36.50% ($p - value < 0.01$).
42, 72	+	Pre-filtering in accordance with the user dimension helped to decrease RMSE by -3.88% in cold-start in case #42 ($p - value < 0.05$) and R^2 increase by 8.85% in case #72 ($p - value < 0.01$).
29, 74, 36, 81	-	IDs-based approaches are poorly performing in cold-start use cases when comparing with the OFFSET strategy's R^2 values (with $p - value < 0.01$).
Timeline tests comparisons with the $OFFSET$ in Table J.5		
8	-	$RMSE$ values are about 4.15% greater in BOOSTER:IDs strategy compared with the $OFFSET$ ($p - value < 0.01$).
1	+	$RMSE$ in FACTORS:IDs is about 1.73% less than in OFFSET ($p-value < 0.01$).
12, 9, 7	+	When using BOOSTER model with ALL, BASE and LOCALITY feature set, we could achieve the smallest $RMSE$ values, from 12.60% to 12.78% less, when comparing with the $OFFSET$ ($p - value < 0.01$).
26, 23	+	We can improve the $NDCG$ metric by 17.94% with $p - value < 0.01$ when using BOOSTER:LOCALITY as compared to the OFFSET strategy in case #26 (versus $NDCG$ improvement by 14.27 % in BOOSTER:BASE in case #23 with $p - value < 0.01$).

Table 9.9 : continued on the next page

Cases	Impact	Findings
39, 35, 38, 41, 37, 40	+	When comparing BOOSTER with the $OFFSET$ strategy, we achieved increased R^2 values by at least 40% when using COUNTRY, ALL, LANGUAGE and DIMENSION feature sets, followed by BASE and LOCALITY with about 39% increase in R^2 (for all mentioned cases, $p - value < 0.01$).
		Timeline tests comparisons with the $BASE$ in Table J.6
1	-	The $RMSE$ values for BOOSTER:IDs were 19.19% greater than BOOSTER:BASE ($p - value < 0.01$).
8	-	The $RMSE$ values for FACTORS:ALL were about 17.91% greater than BOOSTER:BASE ($p - value < 0.01$).
0	+	The BOOSTER:ALL showed about -0.19% smaller $RMSE$ values ($p - value < 0.05$).
18, 14	+	The BOOSTER:LOCALITY and BOOSTER:ALL showed increased NDCG by 3.21% and 2.00% respectively ($p - value < 0.05$).
25, 24, 21, 26, 27	-	The FACTORS model, using COUNTRY, LANGUAGE, BASE, LOCALITY, DIMENSION feature sets, is outperformed by BOOSTER:BASE in the $NDCG$ metric values from 13.52% to 18.79% ($p - value < 0.01$).
31, 28, 30, 33	+	The BOOSTER strategies with COUNTRY, ALL, LANGUAGE ($p - value < 0.01$), and DIMENSION ($p - value < 0.05$) outperform BOOSTER:BASE in R^2 from 1.86 to 0.84%.
34	-	The OFFSET model had lower R^2 performance, by 28.25% with $p-value < 0.01$, than BOOSTER:BASE.
36, 41, 38, 39, 35, 40, 37	-	All models based on factorisation machines showed a lower R^2 performance when compared to BOOSTER:BASE, from 4.54% decrease for FACTORS:ALL in case #36 to 82.34% decrease for FACTORS:IDs in case #37 (all with $p - value < 0.01$).
29	-	The biggest decrease in R^2 of 94.59% was observed for BOOSTER:IDs ($p - value < 0.01$).

Table 9.9: The main findings of the recommendation tests (we specify the positive and negative outcomes in the column "Impact")

9.5.5 Research Questions Revisited

RQ 4.1. Could we find significant differences in user movie genre preferences for different cultural groups in Twitter? The answer is affirmative, we could find statistically significant preferences towards Drama, Action, Animation, Crime and Biography for several cultural groups as shown in appendices I on page 278. Particularly, users tweeted from the inferred Saudi Arabia locations preferred Drama and Action movies when compared to persons from UK and Turkey. Animation was more preferred for persons tweeting from the USA over persons from UK. Overall, Crime was preferred by Multi-Active users when compared with Linear-Active. Considering language locality, Arab speaking users prefer Biography more than English speaking users. It is important to mention that countries and cultural dimensions were inferred, while language preferences were defined in user profiles.

RQ 4.2. Could we improve movie recommendation performance when considering rating locality origins? Overall, $BOOSTER$:LOCALITY and $BOOSTER$:BASE were deemed to be one of the best performing strategies in TIMELINE and user-based offline tests. We observed R^2 performance improvement by 10% using pre-filtering based on user cultural dimension. We achieved 3% improvement in NDCG values for $BOOSTER$:LOCALITY as compared to $BOOSTER$:BASE (case 18 in Table J.6, followed by case 14 for $BOOSTER$:ALL with 2% improvement over the BASE and also having a slight but significant improvement of RMSE values of 0.2% in average in case 0). Interestingly, we also observed a considerable improvement of RMSE by 12% for almost all context inclusion strategies in $BOOSTER$ model as compared with the baseline average-based $OFFSET$ strategy (see cases #12, 9, 7 in Table J.5). The BASE feature set (case #9) was, however, competing with the LOCALITY (case #12), but was outperformed by ALL (case #7).

In cold-start, pre-filtering using user cultural dimension helped to increase R^2 by 8.85% (with $p-value < 0.01$, case #72 in Table J.4). This indicates that selection of one performance metric might be quite misleading when concluding on the recommender performance. In a nutshell, selection of recommendation approaches and the evaluated performance metrics could have an influence on the interpretation of results. It is therefore paramount to consider several recommendation metrics and perform parameters tuning when possible. Since we were able to successfully exploit inferred cultural dimension in pre-filtering user tests, in usual operation and cold-start, we conclude that inferred locality is indeed helpful for better recommendation outcomes to a certain extent, while average based approach could provide a robust baseline especially when used with the pre-filtering step.

9.6 Discussion

9.6.1 Social Media Exploitation and Domain Needs

We investigated the exploitation of pre-filtering and user modeling approaches discussed by Adomavicius and Tuzhilin [7] for creating culture-aware recommendations. As suggested by Hannon with co-authors in [101], user friendship networks could also be exploited for building recommendation strategies. We thus might utilize user networks for mining cultural contexts to complement the lack of information in user profiles and content. However, it is paramount to mention, that particular practical applications could require additional domain-related information as discussed by Swinke [236]. Therefore, a more thoughtful assessment is needed for evaluating culture-aware recommendations in social networking settings when considerable efforts are required to harvest user-related and contextual information useful for particular application goals. As suggested in [5], financial losses and gains should be considered when exploiting additional user or item data in recommender systems. Also, according to Sir Tim Berners-Lee, a possibility of manipulating information contents should not be disregarded in relation with political or commercial gains involved [259].

9.6.2 Movie Genre Preferences

The findings suggest that in some cases (for instance, differences in Drama movie preferences in Arab and Turkish-speaking countries discussed on page 187) we need to be cautious when recommending movies regarding cultural dimensions of users. For particular users, a better recommendation outcome can be achieved by considering their country and language preferences besides other user traits. A further online study is needed to analyse user feedback on received recommendations in different locality settings.

9.6.3 Movie Recommendation Outcomes

We considered the pre-filtering and user modeling approaches as described by Panniello and Gorgoglione [184] while exploiting the contextual information on user cultural origins for creating culture-aware movie recommendation stratagems. Instead of using the information which could be explicitly provided, we infer our contextual data automatically using social media user-generated content. For factoring in user cultural origins (inferred countries, preferred languages, or cultural groups) in our offline user tests, we explored mostly regression models, which are however challenged by robust average-based models not relying on user cultural origins (such as "BASE" model using user and movie numbers with the average movie ratings). We observed that using the gradient boosting model with the "LOCALITY" feature set helped to improve R^2 by 10% over the baseline (case #41). Additionally, we observed a considerable improvement (R^2 increased by 11%, case #72 in Table J.4) while employing pre-filtering based on user cultural dimension and when using average-based baseline.

The overall recommendation performance (in our "timeline" tests) showed even higher increase in R^2 performance when factoring in the inferred country, up to almost 42% over the baseline (case #39 in Table J.5), and the $NDCG$ increase by 17.94% over the baseline (case #26). Even though we achieved a performance improvement while including locality traits into the recommendation models, we must be cautious about selecting predictive model parameters and underlying recommendation technique. Different mixes of contextual parameters lead to different performance outcomes for the selected metrics.

Additionally, we distinguish between different modes of RS operation such as regular operations when user ratings present in the training set, the cold-start problem when there are no previous ratings available. We found out that pre-filtering is quite useful in cold-start situations even with context extracted out of social media. Using user cultural dimensions for pre-filtering, we achieved R^2 increase by 9% (case #72 in Table J.4) and $RMSE$ decrease by 4% (case #42).

9.6.4 Filter Bubble and User Control

Panniello and Gorgoglione [184] showed that in certain e-commerce situations contextual pre-filtering is a feasible approach and might result in better recommendation outcomes considering better precision and lower RMSE, while, however, this could lead to lower recall values when compared with simple collaborative filtering approach without contextual information added. The context is therefore provided explicitly, and users could not have any control over possible missed items due to the exact filtering limitations applied. Users are thus "confined" to the set of preselected items, and this might explain why the recall values drop in the contextual pre-filtering.

When the recommendation system or search engine exploits user context information while limiting information output to specific user characteristics such as location or even particular ideological preferences, a user might be restrained to the predefined information pool while experiencing a lack of "freedom" in information supply. "Filter bubble" (please see [189] referred in [175]) might affect the user experience negatively and even be exploited by political propaganda influencing the behaviour of information systems when possible. When having access to personal data, political parties might utilize information on user preferences and political views to build targeted political messages that could sever not only user privacy right, but their security and democracy could be at stake [259]. In this sense, the content-based recommendation strategies might be more elaborate as compared to the information filtering approach imposed by the collaborative filtering systems such as considering user location information. This is why we might suggest providing a user with the complete control over the filters imposed. Ideally, the information on possible drawbacks of filters usage and thus imposed lack of broad information sources should be available to the user.

Interestingly, Nguyen with co-authors [175] discuss the application of collaborative filtering algorithms for movie recommendations, which overall diversity might be narrowing over time, however, could lead to the more positive user experience for the most actively rating users. Therefore, more research is needed on different recommendation algorithms to study how the recommendation diversity and user satisfaction changes over time in the view of possible filter bubble effects.

9.6.5 Inferring User Locality Traits from Social Media

Throughout this project, we based our culture-awareness and analysis on the assumption that the user origins would match with their locality traits inferred out of user social content and profiles. We might have several reservations about the accuracy of the country/dimension predictive models, which might require further updates with new data published online. User social content might also be misleading. On the other hand, microbloggers might share their posts while travelling to different countries, and therefore their time zones and localities defined in Twitter profiles would not match real user origins.

We tested our country prediction model using the country information provided by Twitter (thus focusing only on the tweets with the country information available). We observed that there are 29 countries predicted with the true positive and negative rates of 75% and above (see Figure H.7 in appendices). When collecting the movie rating dataset, we focused on fifteen countries with the high prediction accuracy of above 90% and most actively rating movies (Table H.2).

However, what if the tweet origins do not correspond with user cultural origins? To find out about country matches between human assessors and geolocations assigned by Twitter, we manually labelled users (who posted their movie ratings with geo-coded tweets) into their countries of origins, and observed a high interrater agreement with Krippendorff's $\alpha = 0.84$ for two human raters and the country labels provided by Twitter (see Table H.4(a)). Persons also might tweet when they live, work or travel in other countries. In accord with GlobalWebIndex research summarised in Twitter infographic [155], 50% of British Twitter users tweet about their holidays, and even more Spanish people use Twitter on holidays tweeting at least once a day while travelling. More than 60% of British persons keep in touch with friends at home using Twitter. Therefore, a persons' social networks could shed light into actual user origins when users travel. The more precise user locality might still be extracted out of social media of the friendship network analysis, and more data can be obtained out of other social networks including Facebook or LinkedIn, potentially providing more finer grained results on user actual locations and also user cultural preferences and origins. The WWW was designed to be a free and open environment for border-less communication, and, social networks could, however, be discouraging sharing free opinions when user privacy and security could be endangered [259].

9.6.6 Movie Ratings Sharing and Self-Image of Users

It is important to think about user intentions of sharing movie ratings and their inherited subjectivity. What is the purpose of sharing movie ratings in Twitter streams, whether it is personal or possibly marketing influence on Twitter followers? Also, some users might also be concerned by their self-image projection while rating movies which might seem to be out of their roles or positions in society. Would a serious adult person be sharing his favourite comedy which would likely appeal more to his son's friends? The possible personal reasons of providing certain ratings cannot be disregarded entirely and thus we might consider more factors than just merely cultural origins, like age, gender, social status, possibly job, personality and other parameters related to users and movies as well.

9.6.7 Application of the Sociological Model

While classifying our users into cultural groups, we based our experiments on the assumptions that the Lewis Model of Cultures is more appropriate to apply to the microblogging communication styles as compared to other models including Hofstede's sociological model. We explain our preference with the focus on communication patterns rather than non-verbal behavioural differences amongst cultures. Even though Lewis model is more flexible in the sense that it does not assume that all persons from particular nations share very similar cultural profiles such as in Hofstede's model, we still stereotype persons into one cultural dimension per country such as in Hofstede's model. We realise that it might have an influence on our results. Notably, for large countries, for instance, the United States of America or Russian Federation, where there are many different cultural groups. In further work, we propose to refine our locations to a higher level of granularity. Some geographic areas might have more diverse populations, while other locations might be more homogeneous. Some locations could match Lewis cultural stereotypes to a larger extent as opposed to others. For instance, we might define Linear-Active cultural dimensions to persons living in Moscow or Saint-Peterborough areas of Russia, or in the East coast of the US. We are aware that languages could provide additional information on person origins. However, we do not want to strictly confine ourselves to stereotyping pitfalls, we instead use the cultural dimension, languages and countries of microposts for user modeling purposes to analyse the strength of the effects of this information when included into recommendation outcomes as an example of "culture-awareness" in adaptive applications such as recommender systems.

9.7 Conclusions and Future Work

One of the primary purposes of recommendation systems is to decrease information overload while providing its users with relevant item suggestions within their context. For instance, in movie recommenders, various factors influence user preferences and needs. User culture is an essential factor, however, less researched in recommender system studies. The integration of culture-related user traits needs more research to find out if this leads to performance gains and improved user satisfaction in particular application areas. With a focus on recommendation performance, we realised several culture-aware recommendation approaches including inferred user origins based on user micro-blogging profiles. In our offline tests, we achieved an improved ranking performance when using added user and movie traits considering locality-specific features including countries, languages and cultural regions based on the existing cultural study. We found user cultural preferences towards movie genres. We also demonstrated challenges of choosing appropriate evaluation metrics. Further research might include online tests to address possible culture-awareness limitations and benefits based on actual user feedback.

The exploitation of the cultural-specific user traits for building culture-aware software applications could provide advantages when the user requires relevant content or functionality, which relate to one's countries or regions. In this respect, we exploited the knowledge on user cultural user origins such as extracted out of user microblogging patterns for building up culture-aware social movie recommendation system. Other possible practical applications could be a friend or local news recommendations. The cultural context of the users could also be used for filtering or ranking recommended items while considering the user cultural traits. In further work, user cultural traits could be exploited for providing more customisation in the context of recommendation diversity. Similarly with the music recommendation diversity needs discussed by Ferwerda with co-authors [69] in relation with user personality traits, we suggest to analyse movie diversity needs in respect to cultural and country-level adaptation contexts.

Part IV

Conclusions

Chapter 10

Conclusions and Further Work

"The best way to predict your future is to create it."

- A. Lincoln, source: goodreads.com

State-of-the-art web applications in e-commerce, e-learning, collaborative platforms for distributed work groups and social networking serve diverse populations of users coming from different places, having their personal preferences and needs for web design, content and functionality. To improve the usability of such web applications and user satisfaction, web developers work on adapting their software to specific, and also culture-related, user needs. For instance, e-learning applications could be further adapted while providing learning material and instruction style according to students' cultural origins [190]. The information overload of users seeking information on the Web might be addressed by applying locality filters. User expectations are not only application dependent, but might also be guided by their personal and cultural preferences. Therefore, we need to create user profiles with information on related user traits and preferences. Such information can be provided explicitly by users, or implicitly extracted from user-generated content, shared opinions and history logs.

We proposed an approach to mine user origins out of social media content, in particular, Twitter microblogs. Previous sociological studies informed us about different culture-specific behaviour patterns in offline communication. We found microblogging patterns in user communication online and discussed culture-specific privacy preferences using openly available microblogging data. Our methodology of creating culture-aware user profiles also enabled us to experiment with movie recommendation strategies, showing a performance improvement in cold-start situations, when user ratings were not yet known to the recommender system. The methodology of mining and exploiting user origins is particularly interesting from a social informatics point of view when technology and society are studied in relation to each other. The described methods use machine learning techniques for profiling microbloggers into cultural groups. This might be interesting for further sociological research aiming at inter-cultural studies in the blogosphere, without approaching persons individually. Instead, for user privacy, such profiling could be performed by aggregating openly-available content without regard to particular users. The profiling outcomes could be used to infer and exploit specific application-specific user preferences. For instance, recommender systems could be provided with user country locations and preferred languages for the benefit of the user.

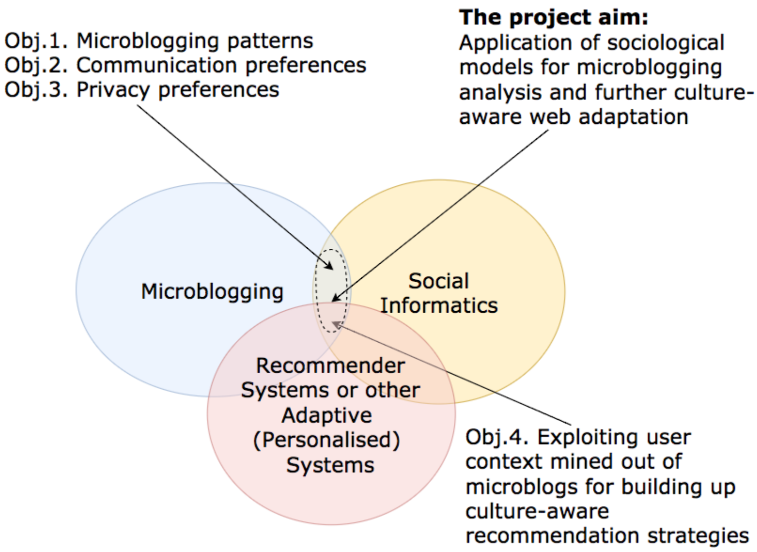


Figure 10.1: Research scope revisited

We can analyse the thesis outcomes at the intersections of several overlapping research fields, as depicted in Figure 10.1, including social informatics, microblogging and recommender system design.

Objectives	Achieved results
Goal 1: How cultural differences impact user behaviour in Twitter?	
Objective 1	Determined microblogging patterns in terms of Twitter feature usage including URLS, Hashtags, User Mentions, Replies and Days of Microposting
Objective 2	Found important features for analysis of communication preferences in respect to their cultural origins
Goal 2: How to exploit user microblogging patterns in further web adaptation?	
Objective 3	Using microblogging patterns for user origin prediction enabled (to a certain extent) identification of microposting country and region origins, and analysis of user privacy preferences
Objective 4	Exploited user locality context inferred from microblogging user profiles enabled creation of culture-aware recommendation strategies outperforming average-based baseline recommendations

Table 10.1: Project objectives achieved

Research Question	Answer	Pages
Objective 1: Finding Cultural Cues Online, addressed in [121]		
RQ 1.1: Could we find differences in Twitter features usage for persons microposting from different geographic regions (origins)?	Yes, we found significant differences of Twitter features usage for five countries in Twitter (Japan, Spain, Germany, the US and Brazil) with $p < .001$ significance level using t-tests.	ch. 6, pp. 100
RQ 1.2: Could we exploit these differences for predicting user origins on a country and geographic region level?	Yes, user behavioural differences in Twitter enabled us to achieve 10-fold cross-validation error of 0.42 and 0.29 for Country and Cultural region levels. When adding the user language, the cross-validation error decreased respectively to 0.06 and 0.04.	ch. 6, pp. 106
Objective 2: Determining microblogging features importance for predicting user origins and communication preferences [52]		

Table 10.2: continued on the next page

Research Question	Answer	Pages
RQ 2.1: Could we exploit user contact network, i.e., friends, for predicting user cultural origins?	Yes. Although we could not exploit user contacts' network meta-data match for inferring user country locations, however, FOLLOWERS feature set extracted from the user followers network enabled us to infer the user country locations with a 3-times cross-validation test results accuracy of more than 84%, when FOLLOWERS features set was combined with user language we achieved 94% of accuracy for users having top most used language in respective country.	ch. 7, pp. 127
RQ 2.2: What microblogging features (user-related and friend-related) are the most prominent in revealing user cultural traits in Twitter?	We found that user language, followers' majority country, and time zones are the most revealing user country locations based on the importance of the features extracted from the decision trees.	ch. 7, pp. 129
RQ 2.3: Could we find communication preferences in respect to user cultural origin?	Yes. We found that cultural dimension match is associated (statistically significant) with 23% higher probability of user responds when analysing statistics on logit marginal effects, and found that users tend to respond mostly within their cultural regions.	ch. 7, pp. 133, 130
Objective 3: Monitoring Twitter privacy settings usage [53]		
RQ 3.1: Are there differences in privacy settings usage by different cultural groups?	Yes, Multi-active users group has a larger fraction of open profiles, while Reactive user group has a greater fraction of closed profiles.	ch. 8, pp. 150

Table 10.2: continued on the next page

Research Question	Answer	Pages
RQ 3.2: How does protecting user accounts affect user communication styles in Twitter?	Interestingly, protected user profiles have a smaller contacts network and less listed, however, not less influential compared with open user profiles. The higher number of setting changes characterises protected user profiles. They also favourite other user messages more frequently.	ch. 8, pp. 148
RQ 3.3: How could user privacy preferences be exploited in real-life scenarios (discussion on security implications and related issues)?	User locations could be leaked via user friendship networks, or from user-generated content and meta-data when user profiles are open. Users should be provided with full control over their content and profile storage. The privacy needs differ across the cultures.	ch. 8
Objective 4: Creating culture-aware social recommender		
RQ 4.1: Could we find statistically significant movie genre preferences in relation to user inferred origins?	We found statistically significant genre preferences for many locality groups, including different country, language, and cultural dimension groups. For instance, users from Saudi Arabia prefer Drama and Action movies more (give higher ratings in average) compared to users from the UK and Turkey.	ch. 9, pp. 278

Table 10.2: continued on the next page

Research Question	Answer	Pages
RQ 4.2: Could we improve movie recommendation performance when considering user origins and other item or user-related features?	Yes, to a certain extent, in a cold-start situation when pre-filtering on user cultural dimension while using average-based OFFSET model. However, the offline user tests analysis showed that user modeling was not successful as compared to pre-filtering using user cultural dimension helping us to increase R^2 by 11% in usual conditions (case 42) and by 9% in the cold-start situation (case 72) for average-based recommendation strategy. In timeline tests, usage of the LOCALITY feature set enabled improvement $NDCG@10$ metric values of the BOOSTER user modeling compared to BOOSTER BASE by 2.7% (case 18). Generally, we achieved slightly better recommendation performance when adding ALL user and item related features, inferred country, or when using LOCALITY feature set based on recommendation strategy average ranks in timeline tests.	ch. 9, pp. 313, 313, 193

Table 10.2: Research questions answered

Table 10.1 outlines the the achieved research objectives and Table 10.2 provides the answers to the related questions.

Since user locations on Twitter are not available or inaccurate in the majority of cases, we created user models which were further used for predicting user origins (addressing Objective 1, in which the patterns found enabled us to distinguish between cultural and country groups, chapter 6 “User Origin Prediction”). For this, we used Twitter-specific features extracted from users’ content and social networks. With the information on user origins, we analysed user communication (Objective

2, in which we observed that microbloggers tend to respond the most within their cultural groups, chapter 7 "Communication Preferences") and privacy preferences in Twitter (Objective 3, our study of Twitter settings usage in chapter 8 "Privacy Settings Usage in Twitter" revealed the most "private" cultural group in Twitter). We assessed different recommendation approaches (Objective 4, chapter 9 "Culture-aware Social Recommenders") considering cultural contexts of the users. With this, we partially (in the case when culture-awareness is advantageous in the recommender system application) addressed the cold-start problem, which is widely discussed in recommender systems research, and lesser-known culture-aware adaptation. Further, we critically discuss the contributions mentioned above, their standing in the light of previous research, applicability in practice, research implications, limitations of proposed solutions and conclude with further work insights.

10.1 Microblogging Behaviour Patterns

The main emphasis of the current Web is on Social Networking and as a consequence, large volumes of user-generated content could be analysed for a better understanding of user behaviour online and their preferences. The prominent research works inform us of cultural differences of exploiting SN services such as microblogs [80, 225, 180]. Based on the insights of previous work and sociological studies [148, 149, 112], we assumed that persons from different origins might also have different online behaviour patterns. Our contributions further summarised helped us to confirm this assumption.

In chapter 6 "User Origin Prediction", we analysed how users from different cultural origins employ Twitter. Interestingly, our results correspond with some findings of Poblete with co-authors [193], in which users from the USA lead in Uniform Resource Locator (URL) sharing. Similar user behaviour was observed in [121, 119], in which users from the USA and Germany tend to share URLs and Twitter hashtags the most compared to other user groups analysed. The main findings reveal a culture-specific user behaviour on Twitter discussed regarding previous sociological research by Lewis [148]. Since the results enable us to distinguish between different cultural origins of user groups, a culture-oriented user modeling approach was proposed based on the analysis of microblogging behavioural patterns in [121].

From a practical point of view, we might exploit user online activities for predicting user origins and adapting web applications when culture-awareness is beneficial. For instance, e-commerce websites could provide their users with goods or content tailored to their cultural backgrounds. We could also design social networking sites with user culture in mind. For instance, a tagging functionality could be appreciated by users from Germany, while reply functionality would be more accessible to Japanese users. Further work might be performed on whether such design adaptations would be met with better user satisfaction. Further, we based the research on a microblogs analysis. This limitation could be addressed by applying the methodology described using other social media sources where user profiles are openly available. A longitudinal study could also inform us whether culture-specific microblogging patterns change over extended periods of time. However, we should refer to the previous research by Soliman with co-authors [229] indicating that user tweeting activities usually reduced within few months.

10.2 User Origins Prediction

In our experiments, less than 2% of tweets were geographically tagged. Nevertheless, microblogging behaviour patterns enabled us to exploit Twitter-specific features for building up country and cultural groups predicting models based on Decision Trees. Even though our best performing country prediction strategies achieved an accuracy of more than 90%, other feature sets and classification methods such as Support Vector Machines could be considered.

Compared with other related works, for instance, by Alex [13] achieving 90% of precision in user city location prediction for several countries, our location detection technique is not finely-grained, and we aim to predict locations on a country-level only. We work however with many nations and languages while combining tweet location and profile location data together with the timezone and user languages defined in the Twitter profile. This enables us to avoid reliance on gazetteers, and also external services such as Yahoo geo-parser exploited by Kulshrestha in [142] while achieving a reasonable country location detection without extensive resource consumption when building on tweets with geographic location enabled for training and testing purposes. When considering only country-level predicting, we outperformed the "Calgali" algorithm by Hecht [104] which uses Twitter contents as compared to our approach of considering follower networks and languages defined in user profiles, as described in chapter 7 "Communication Preferences".

We agree with Alex [13] that there are more application possibilities not restricted to sentiment analysis and information flows. Using user country inference could help in more general tasks like adapting e-commerce and e-learning and while building distributed international teams when information on user cultural origins is paramount for the application goals.

10.3 Communication Preferences

The information on inferred user countries allowed us to study user communication in Twitter. In chapter 7 "Communication Preferences", we analysed microblogs of users from the top 13 most active countries on Twitter. We exploited Tweets' content and metadata as features for building a classification model for predicting user countries. The results showed that using user-related microblogging features and features extracted from a followers' network enabled the prediction of user countries with an accuracy of more than 90%. The most important features included user languages, countries and time zones of their followers, followers influence (the number of followers divided by the sum of followers and friends), and microblogging days of the week. While mining user country locations might be useful, the privacy concerns should be observed. Our approach showed a good efficiency for detecting tweet origins. This information could, however, be abused in situations when user whereabouts should be protected.

Having knowledge of user origins, we analysed user replies to their friends. Our findings showed that Twitter users were most likely to reply to friends coming from the same region; however, country matches were less important for getting user replies. Future work might reassess the social networking and related cultural models in view of online communication preferences since a person tends to respond more actively within their cultural group. We might hypothesise that creating more nationally diverse teams can lead to greater communication when their (team) members are within similar cultural regions. We proposed to further exploit these cultural regions in recommender systems (for instance, friend or collaborator suggestions), which require a user cultural context. We also discussed the subject of web application adaptation in the scope of preserving online privacy.

10.4 Privacy Needs

Even though we work with aggregated data for building user profiles, and do not operate with individual sensitive data, the possibility of exploiting user contextual information, for instance, out of social networks, could not be disregarded as we suggest in chapter 7 "Communication Preferences". Further in chapter 8 "Privacy Settings Usage in Twitter", we analysed user microblogging preferences in respect to privacy considerations, which are paramount when personal and sensitive information could be leaked out of publicly available microblogs, for instance, via friendship networks. Our findings revealed that a relatively small fraction (about 5%) of our followed users still require protecting their profiles. Interestingly, "Reactive" users from Indonesia and Japan have the largest fractions of protected profiles compared to other cultural groups analysed. We propose to further analyse culture and country-specific privacy preferences in other social networking platforms and web applications. Privacy settings in social networks and default user profiles might be defined with user cultural needs in mind. Additionally, users' complete control over shared data is needed.

In this thesis, we focused mostly on the culture-awareness for web applications that adapt to user origins. We must be aware however that in the movie recommendations or other personalised services, the culture of users is not the only factor to consider. In the practical applications, we want to analyse the benefits and possible drawbacks of using different features, for instance, personal preferences, tastes, moods and even personality traits. For example, a study by Chorley et al. [41] on personality and visited places on Foursquare informs us about a positive relationship between usage of location-based social networks and conscientiousness. In this case, we might re-evaluate the cases of privacy needs for persons willing the share their geographic locations as opposing to persons who do not share their places in social networks (referring to chapter chapter 8 "Privacy Settings Usage in Twitter"). Noë et al. [176] discuss the correlation between personality traits and network-based features extracted from Facebook data, they found that extroversion and social network size are positively correlated. The future research might further investigate the possible scenarios of mixing culture-related factors and personality traits to study their relative impact on recommendations or web search outcomes.

10.5 Culture-awareness Using Inferred Origins

Finally, we demonstrated the exploitation of inferred locality traits in a movie recommendation domain. While adding inferred locality traits, we improved recommendation performance when compared with the average-based baseline and also non-locality-aware recommendation strategies. The results are especially pertinent to cold-start situations when users are new to the system, and no previous history of ratings is available. Using pre-filtering on user cultural regions and factoring user locality traits into our user modeling approach enabled us to improve movie rating predictive performance (R^2 particularly) by at least 10%. We also observed significant differences in movie genre preferences amongst the analysed inferred locations. We are however aware of possible limitations due to our specific dataset retrieved with the Twitter API. We also cannot ensure that the offline tests could be used to reflect user satisfaction in reality. This is why we further plan to perform online tests to get real-life user feedback on the proposed culture-aware recommendations. Online tests would enable us to additionally evaluate our locality prediction classification model based on metadata extracted from Twitter profiles.The movie recommendation domain is not the only potential application of culture-awareness.

Since 2010, we observe that research in the area of culture-centered user interface design is rapidly evolving [106]. E-commerce and e-learning domains might benefit from knowledge on user cultural origins. Learning styles and e-learning environment preferences also differ for persons from particular cultural backgrounds. Social networks could be used to obtain such information unobtrusively. However, such information gathering should be performed respecting user privacy and full user control over the accessed data. We emphasised online user privacy issues in respect of existing online privacy guidelines and challenges to date. We also discussed dealing with sensitive user information while doing social networking research. We proposed to work with anonymous and aggregated user context instead of dealing with finer user data such as precise user locations. The privacy and security threats should not be underestimated when thinking about human life reflected in social media. Nevertheless, the possibility of mining cultural user traits and preferences provides immense application potential. Further research might go more in-depth into an evaluation of the financial benefits and overall user satisfaction with adapted systems, recommenders or other intelligent state-of-the-art web applications benefiting from knowledge on person origins.

10.6 Research Scope Limitations

It is important to mention that the project scope did not include other exciting directions such as sentiment analysis of the social media content and usage of different machine techniques, for instance, clustering, for grouping users into their respective cultural groups automatically. We intentionally excluded these topics for several reasons.

First of all, our objectives were designed to demonstrate a possible solution for inferring user origins for further web adaptation. Twitter microblogs were used for training our models in a semi-supervised way. We have considered the geolocations attached in the metadata of user-generated content. First of all, the machine learning approach employed was based on decision trees and using geolocations as labels might provide a precise solution. We, however, performed manual labeling to test our automatic approach, with about 66% agreement between human raters and the automatic classification, and about 78% agreement between human raters and Twitter geolocations (on a small sample of 56 users actively tweeting on movie ratings and also having geolocations attached). We are aware of this limitation and propose to extend our approach to improve the accuracy of user location inference.

Clustering methods could also be considered in further work which focuses on the defined number of user cultural groups. Our approach, however, aims at a flexible way to infer user country locations without looking into the context. This allows us to be less dependent on translation services and the external assistance of human translators to do content analysis, which is prone to human errors. This research, however, might benefit from more in-depth content analysis of different cultural groups while building on the previous research work by [40] (exploring tweets usage for geolocating users on a city level), [226] (websites content analysis revealed the influence of cultural values) and the country inference solution proposed in this thesis. The further work could extend the proposed approach for detecting user locations on a city level while analysing culture-specific content features. For instance, different cultural groups might have their own online content preferences and marked online products.

Moreover, analysis of the user-generated content published in borderless communication requires also language-specific considerations. For instance, the polarity phrases and related words in sentiment analysis require to have related language expertise. The meaning of the free-text could also be changed due to different semantics attached by the content publishers and their emotional state or another context. Therefore, more work needs to be done in the sentiment analysis field within cultural online behaviour research. We can learn from the previous work [208] on multilingual sentiment analysis using classification techniques while analysing the distribution of sentiment values for several cities and languages.

Cultural sociology could benefit from big data sources and their state-of-the-art analysis techniques [18]. Specialised search engines and social media analysis tools such as provided by LexisNexis, application program interfaces for access to big data, automated text classification and clustering for topic extraction techniques, topic modeling, and other tools could benefit from big data streams, which can be supplied with geolocation metadata [18]. Our solution was built on top of the social media content, and we are grateful for the thousands of Twitter users for their contributions.

However, the exploitation of the existing computational techniques remains challenging for sociological researchers [18]. The lack of technical knowledge might pose a difficulty while accessing and mining big data of user-generated context. Secondly, the data might be quite unstructured and lack user context. The free-text could lack semantics, and we cannot access non-verbal cues coming from individual users [18]. Therefore, we propose to further extend our approach with sensory information, for instance, provided with the help of ubiquitous computing techniques, web cameras, voice recognition, and other modalities helping to understand user context, psychological and emotional traits.

10.7 Application of Machine-learning Algorithms

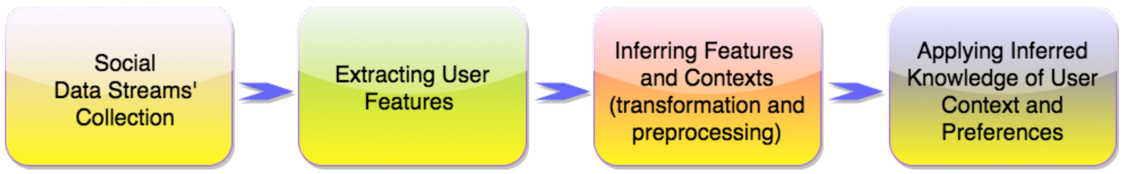


Figure 10.2: Learning user preferences and context from social streams

In this thesis, Twitter streams of user-generated context and associated metadata were used to extract application-specific features while addressing particular research questions. Figure 10.2 shows that we infer user traits out of social data while preprocessing the extracted features.

We exploited several machine learning techniques such as decision trees, gradient boosting regression and factor machines. For our research goals, this “offline learning” approach was sufficient. However, further research could benefit from the application and assessment of incremental or “online learning”. We might consider exploiting neural network-based algorithms for incremental learning in the context of personalised solutions such as described in [77] enabling us to learn user preferences rapidly in real-time. However, caution is required when designing algorithms without previous knowledge of the dataset set when learning online. The feasibility, performance, and scalability of the algorithm's adaptation to online learning are yet to be investigated. This becomes very important due to the rapidly increasing user-generated content over time, and also the development of decentralised cloud-based platforms potentially facilitating research in the social studies and human-computer interaction fields.

10.8 Summary

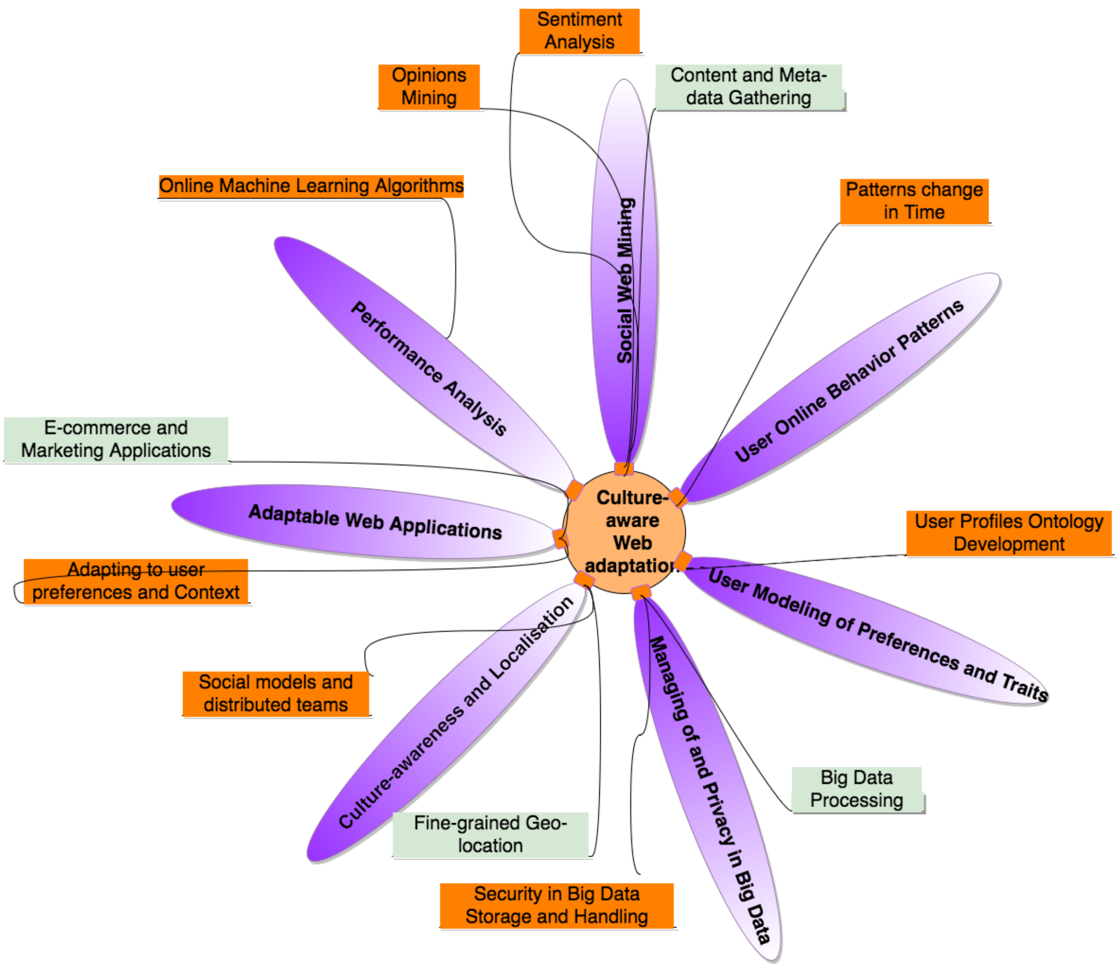


Figure 10.3: Future work directions (the flower chart depicts the research issues discussed throughout thesis as flower petals, its stamens point into further research work in (coloured in orange) and related engineering problems (coloured light green)

To summarise, we applied the existing sociological model of offline communication to online communication in microblogs. We found Twitter microblogging patterns allowing us to infer user origins, and exploit this information for investigating microblogger communication preferences with their networking friends and privacy settings usage while addressing the cold-start problem in a movie recommender. We could not find an explicit mapping between the Lewis model of cultures and microblogging behaviour. However, we found notable behaviour differences between different cultural groups (particularly Reactive users from Japan). Our methodology is a possible solution to perform user locality inference for further user profiling and creating adaptive applications when user origins are paramount. While being restricted to Twitter microblogs and within a set of analysed countries, we believe that our approach could be extended to investigate different social media sources while building on the current sociological research with new findings on user behaviour online. The full extent of the financial and societal implications will be assessed in future research. Overall, we created a set of solutions and discussed further research insights outlined in Figure 10.3 including sentiment and opinion analysis of data streams coming from different inferred locations, social web behaviour changes in time, scalability and applications of machine learning algorithms for big datasets for various applications in web research and e-commerce. We also recommend further adapting and assessing the proposed approach for the application of fast online incremental learning algorithms in the personalisation domain and adaptive social applications while benefiting from modern cloud-based solutions.

Part V

Appendices

Appendix A

Twitter Features Usage Among the Country Groups

We performed a two-step analysis to identify significant difference for each Twitter feature usage among the selected country groups in chapter 6 “User origin Prediction”. First, the one-way Analysis of Variance (ANOVA) was performed to find out that there are statistically significant differences between the means of features analysed as shown in Table A.1. For each of the ten analysed features, we tested the null hypothesis Ho, which was rejected when at least two group means were significantly different from each other. The ANOVA F-value was estimated with 5% significance level for all features analysed. To summarise, we concluded that there is a significant difference among the country groups for all features analysed with $p < 0.05$. Due to normality and variances equality assumptions violations, we performed features transformation onto their square roots and report also non-parametric Kruskal-Wallis test results in Table A.2. Secondly, Table A.3 shows the two-sample Welchs t-test[1] results on the country level assuming non-equal variances (please refer to 98 for justification of the selected test).

[1]Using MATLAB two-sample t-test (ttest2 with ‘unequal’ option for ‘Vartype’ parameter)

	Sum of Squares	df	Mean Square	F	Sig.
Mobility: Number of Different Countries Detected in Tweets minus one					
Between Groups	177.39	4	44.35	259.12	$p < 0.01$
Within Groups	2052.57	11993	0.17		
Total	2229.96	11997			
Followers: Number of Followers					
Between Groups	19686.92	4	4921.73	$p < 0.01$	
Within Groups	1241088.9	11993	103.48		
Total	1260775.82	11997			
Friends: Number of Friends					
Between Groups	8353.19	4	2088.3	28.24	$p < 0.01$
Within Groups	886927.03	11993	73.95		
Total	895280.21	11997			
Languages: Number of Languages detected in Tweets minus one					
Between Groups	1546.93	4	386.73	1653.67	$p < 0.01$
Within Groups	2804.71	11993	0.23		
Total	4351.64	11997			
Tags: Number of Hashtags found in Tweets					
Between Groups	19876.81	4	4969.2	956.05	$p < 0.01$
Within Groups	62335.27	11993	5.2		
Total	82212.07	11997			
Tags: Number of Web Links found in Tweets					
Between Groups	1669.86	4	417.46	73.96	$p < 0.01$
Within Groups	67693.53	11993	5.64		
Total	69363.39	11997			
Reply: Number of Tweets with Replies					
Between Groups	865.34	4	216.34	53.01	$p < 0.01$
Within Groups	48944.42	11993	4.08		
Total	49809.76	11997			
Retweets: Number of Tweets with Retweets					
Between Groups	7702.69	4	1925.67	612.12	$p < 0.01$
Within Groups	37729.13	11993	3.15		
Total	45431.82	11997			

Table A.1: continued on the next page

	Sum of Squares	df	Mean Square	F	Sig.
Mentions: Number of User Mentions found in Tweets					
Between Groups	10685.48	4	2671.37	550.31	$p < 0.01$
Within Groups	58218.1	11993	4.85		
Total	68903.57	11997			
Weekends: Number of Tweets posted on Weekends					
Between Groups	474.02	4	118.51	115.26	$p < 0.01$
Within Groups	12330.41	11993	1.03		
Total	12804.43	11997			

Table A.1: ANOVA results for country-level comparison of Twitter feature usage

Feature	H statistic	p-value
Mobility: Number of Different Countries Detected in Tweets minus one	538.98	$p < 0.01$
Followers: Number of Followers	193.23	$p < 0.01$
Friends: Number of Friends	179.36	$p < 0.01$
Languages: Number of Languages detected in Tweets minus one	3578.18	$p < 0.01$
Tags: Number of Hashtags found in Tweets	3545.52	$p < 0.01$
Tags: Number of Web Links found in Tweets	317.12	$p < 0.01$
Reply: Number of Tweets with Replies	208.97	$p < 0.01$
Retweets: Number of Tweets with Retweets	2154.56	$p < 0.01$
Mentions: Number of User Mentions found in Tweets	2048.17	$p < 0.01$
Weekends: Number of Tweets posted on Weekends	434.71	$p < 0.01$

Table A.2: Kruskal-Wallis test results for country-level comparison

G_1	G_2	t	df	p	μ_1	μ_2	G_1	G_2	t	df	p	μ_1	μ_2
RQ1: Hashtags													
ES	BR	22.82	6500.3	< 0.001	29.56	14.76	**ES**	**US**	0.88	2192.8	> 0.05	29.56	28.67
ES	JP	38.15	5755.4	< 0.001	29.56	7.63	ES	DE	-4.89	2942.4	< 0.001	29.56	34.41
BR	US	-14.36	1973.6	< 0.001	14.76	28.67	BR	JP	13.64	5713.1	< 0.001	14.76	7.63
BR	DE	-20.49	2655.8	< 0.001	14.76	34.40	US	JP	22.84	1642.9	< 0.001	28.67	7.63
US	DE	-4.68	3005.1	< 0.001	28.67	34.41	JP	DE	-29.40	2217.7	< 0.001	7.63	34.41
RQ2: URLs													
ES	**BR**	-1.93	6069	> 0.05	30.78	32.09	ES	US	-13.96	2279.9	< 0.001	30.78	42.46
ES	JP	-2.15	6917.4	< 0.05	30.78	32.09	ES	DE	-8.64	3287	< 0.001	30.78	37.49
BR	US	-11.55	2798.2	< 0.001	32.09	42.46	**BR**	**JP**	0.001	6193.2	> 0.05	32.09	32.09
BR	DE	-6.40	3934.9	< 0.001	32.09	37.49	US	JP	12.27	2356.8	< 0.001	42.46	32.09
US	DE	5.12	2900.6	< 0.001	42.46	37.49	JP	DE	-6.86	3398	< 0.001	32.09	37.49
RQ3: Languages													
ES	BR	-4.45	5930	< 0.001	1.067	1.16	ES	US	39.78	4235.4	< 0.001	1.066	0.20
ES	JP	48.54	4399.1	< 0.001	1.06	0.16	ES	DE	3.01	4346.7	< 0.01	1.06	0.993
BR	US	48.89	3732.2	< 0.001	1.16	0.20	BR	JP	61.85	4589.7	< 0.001	1.16	0.16
BR	DE	7.76	3781.7	< 0.001	1.16	0.99	US	JP	2.55	JP3	< 0.05	0.20	0.16
US	DE	-36.85	2937.6	< 0.001	0.20	0.99	JP	DE	-45.30	2394.4	< 0.001	0.16	0.99
RQ4: Mobility													

Table A.3: continued on the next page

G_1	G_2	t	df	p	μ_1	μ_2	G_1	G_2	t	df	p	μ_1	μ_2
ES	BR	-2.96	6464.5	< 0.05	0.87	0.90	ES	US	-3.26	2610.1	< 0.05	0.87	0.91
ES	JP	22.08	6907.9	< 0.001	0.87	0.601	ES	DE	-6.89	2942.5	< 0.001	0.87	0.98
BR	**US**	-1.13	2234.8	> 0.05	0.90	0.91	BR	JP	26.59	6412.7	< 0.001	0.90	0.60
BR	DE	-5.16	2606.1	< 0.001	0.90	0.98	US	JP	20.99	2755.1	< 0.001	0.91	0.60
US	DE	-3.56	3080.7	< 0.001	0.91	0.98	JP	DE	-22.47	3073.7	< 0.001	0.60	0.98
RQ5: Weekends													
ES	**BR**	-1.28	6453.4	> 0.05	24.04	24.35	**ES**	**US**	1.81	2576.8	> 0.05	24.04	23.50
ES	JP	-19.34	6924.7	< 0.001	24.04	28.57	ES	DE	-4.3	3577.3	< 0.001	24.04	25.26
BR	US	2.79	2719.3	< 0.05	24.35	23.53	BR	JP	-17.41	6473.5	< 0.001	24.35	28.57
BR	DE	-3.15	3727.2	< 0.01	24.35	25.26	US	JP	-17.09	2573.1	< 0.001	23.50	28.57
US	DE	-5.23	2949.2	< 0.001	23.50	25.26	JP	DE	11.63	3574.2	< 0.001	28.57	25.26
RQ6: Friends													
ES	BR	6.01	6545	< 0.001	335.37	282.14	ES	US	-4.69	2074.3	< 0.001	335.37	400.5
ES	**JP**	-0.20	6524.7	> 0.05	335.37	337.5	**ES**	**DE**	-1.70	3138.6	> 0.05	335.37	356.2
BR	US	-8.63	1987.4	< 0.001	282.14	400.5	BR	JP	-5.36	6200.2	< 0.001	282.14	337.5
BR	DE	-6.13	2981.4	< 0.001	282.14	356.2	US	JP	4.25	2591	< 0.001	400.55	337.5
US	DE	2.75	2807.3	< 0.01	400.55	356.2	**JP**	**DE**	-1.41	3930.4	> 0.05	337.47	356.2
RQ7: Followers													
ES	BR	-3.13	6398.3	< 0.01	296.45	335.8	ES	US	-9.19	1771.3	< 0.001	296.45	501.6

Table A.3: continued on the next page

G_1	G_2	t	df	p	μ_1	μ_2	G_1	G_2	t	df	p	μ_1	μ_2
ES	**JP**	-1.82	6925	> 0.05	296.45	318.3	ES	DE	-6.27	3078	< 0.001	296.45	398.9
BR	US	-7.31	1870.1	< 0.001	335.81	501.6	**BR**	**JP**	1.39	6425.6	> 0.05	335.81	318.3
BR	DE	-3.75	3308.2	< 0.001	335.81	398.9	US	JP	8.21	1774.2	< 0.001	501.6	318.3
US	DE	4.11	2421	< 0.001	501.6	398.9	JP	DE	-4.93	3087.1	< 0.001	318.27	398.9
RQ8: User Mentions													
ES	BR	30.23	6520.2	< 0.001	83.91	57.94	ES	US	7.18	2215.8	< 0.001	83.912	75.13
ES	JP	48.54	6555.9	< 0.001	83.91	46.51	ES	DE	17.57	3613.8	< 0.001	83.91	65.83
BR	US	-14.02	2221	< 0.001	57.94	75.13	BR	JP	14.76	6036.8	< 0.001	57.94	46.51
BR	DE	-7.65	3588	< 0.001	57.94	65.83	US	JP	24.56	1869.7	< 0.001	75.13	46.516
US	DE	6.88	2674.4	< 0.001	75.13	65.83	JP	DE	-20.13	2960.3	< 0.001	46.516	65.83
RQ9: Replies													
ES	BR	12.63	6447.3	< 0.001	27.47	22.04	ES	US	2.20	2242	< 0.05	27.474	26.18
ES	**JP**	0.50	6673.2	> 0.05	27.47	27.24	ES	DE	-2.11	3111.2	< 0.05	27.47	28.66
BR	US	-6.92	2367.5	< 0.001	22.04	26.18	BR	JP	-10.89	6540.9	< 0.001	22.04	27.24
BR	DE	-11.53	3266.8	< 0.001	22.04	28.66	**US**	**JP**	-1.70	2698.7	> 0.05	26.18	27.24
US	DE	-3.55	2959.1	< 0.001	26.18	28.66	JP	DE	-2.37	3705.3	< 0.05	27.24	28.66
RQ10: Retweets													
ES	BR	24.73	6475.3	< 0.001	23.16	14.26	ES	US	17.77	2832	< 0.001	23.16	15.00
ES	JP	44.29	6278.1	< 0.001	23.16	8.22	ES	DE	18.89	3973.3	< 0.001	23.16	14.95

Table A.3: continued on the next page

G_1	G_2	t	df	p	μ_1	μ_2	G_1	G_2	t	df	p	μ_1	μ_2
BR	**US**	-1.68	2415.4	> 0.05	14.26	15.0	BR	JP	19.76	6256.1	< 0.001	14.26	8.22
BR	**DE**	-1.66	3418.3	> 0.05	14.26	14.95	US	JP	16.19	2107.9	< 0.001	15.00	8.22
US	**DE**	0.11	2919	> 0.05	15.0	14.95	JP	DE	-17.18	3001.8	< 0.001	8.22	14.95

Table A.3: Comparing Twitter features usage for country groups using Welch's t-test (rows shown in bold font indicate the cases with no significant differences between group means and $p > 0.05$)

Appendix B

Flight and Cultural Score Distances between Countries

Tables B.1 and B.2 are used to draw Figure 4.2 in chapter 4 "Models of Culture". Scores for Geert-Jan Hofstede's model provided by Markus [158], scores for cultural dimensions of The Lewis Model were derived from the approximate distances to the apexes of the triangle [45], scores for model by Fons Trompenaars and Charles Hampden-Turner were assigned from the levels of "low", "middle" and "high", further converted into numerical values, based on country lists [165, 166].

	BR	GB	GR	IN	ID	IT	JP	RU	SA	ES	SE	TH	TR	US	VE
GB	5559	0													
GR	5994	1621	0												
IN	9188	4767	3596	0											
ID	11023	7314	6368	2786	0										
IT	5628	1181	524	4087	6841	0									
JP	10794	5732	5861	3708	2988	6059	0								
RU	8978	3501	3700	3096	4319	3830	2257	0							
SA	7049	3275	1717	2177	4913	2239	5438	3810	0						
ES	4868	1034	1358	4941	7690	853	6633	4376	3034	0					
SE	6418	876	1461	3982	6447	1288	4968	2712	2807	1667	0				
TH	2682	5873	4966	1483	1446	5420	2682	3157	3658	6261	5013	0			
TR	6682	2106	721	2881	5648	1209	5311	3271	1189	2062	1631	4245	0		
US	4537	4255	5860	8448	9303	5369	6321	5536	7525	4730	4776	8630	6341	0	
VE	1740	4814	5859	9454	12042	5368	9088	7728	7392	4515	5671	10682	6574	2797	0

Table B.1: Flight distances in miles between countries

(with distancefromto.net using Google Maps)

	BR	GB	GR	IN	ID	IT	JP	RU	SA	ES	SE	TH	TR	US	VE
GH_ID	0.33	0.97	0.29	0.46	0.03	0.81	0.43	0.34	0.16	0.49	0.75	0.10	0.32	1.00	0.00
GH_LI	0.49	0.61	0.38	0.07	0.23	0.12	0.28	0.00	0.40	0.30	0.72	0.31	0.36	0.60	1.00
GH_LT	0.39	0.49	0.40	0.49	0.64	0.62	1.00	0.90	0.28	0.44	0.51	0.22	0.42	0.14	0.00
GH_ML	0.49	0.68	0.58	0.57	0.46	0.72	1.00	0.34	0.61	0.41	0.00	0.32	0.44	0.63	0.76
GH_PD	0.59	0.06	0.45	0.72	0.73	0.30	0.36	0.97	1.00	0.41	0.00	0.52	0.55	0.14	0.78
GH_UA	0.66	0.08	1.00	0.15	0.27	0.65	0.89	0.93	0.72	0.80	0.00	0.49	0.79	0.24	0.66
FT_AC	0.50	0.50	0.50	0.50	0.00	0.50	0.50	0.50	0.00	0.50	0.50	0.50	0.50	1.00	0.50
FT_ID	0.50	1.00	0.50	0.50	0.50	0.50	0.00	0.50	0.50	0.50	1.00	0.00	0.50	1.00	0.00
FT_IN	0.50	1.00	0.50	0.00	0.50	0.50	0.50	0.00	0.00	0.50	0.00	0.50	0.50	1.00	0.50
FT_NE	0.00	1.00	0.50	0.50	0.50	0.00	1.00	0.50	0.50	0.00	0.50	0.50	0.50	0.50	0.00
FT_SP	0.50	1.00	0.50	0.00	0.50	0.50	0.00	0.00	0.50	0.00	1.00	0.50	0.50	1.00	0.00
FT_ST	0.50	1.00	0.50	0.50	0.50	0.50	0.00	0.50	0.50	0.50	1.00	0.50	0.50	1.00	0.50
FT_UN	0.50	1.00	0.50	0.50	0.00	0.50	0.00	0.00	0.50	0.50	1.00	0.50	0.50	1.00	0.00
RL_LA	0.00	1.00	0.00	0.00	0.00	0.00	0.14	0.29	0.00	0.00	0.86	0.00	0.00	1.00	0.00
RL_MA	1.00	0.00	0.88	0.50	0.38	0.88	0.00	0.75	0.75	0.88	0.00	0.25	0.62	0.12	1.00
RL_RE	0.00	0.14	0.00	0.57	0.71	0.00	1.00	0.00	0.29	0.00	0.29	0.86	0.43	0.00	0.00

Table B.2: Cultural dimension scores for countries (normalised)

Appendix C

Country Prediction using Twitter

C.1 Preliminary Tests

Herein we performed preliminary tests to assess the feasibility of tweet origin country prediction using several clarification models. The feature set includes BEHAVIOUR (described in "Experimental Setup" section in ch.6 on page 94) features. The classifiers were created on 11998 user profiles having geo-tagging information directing to five countries: Germany, Brazil, Japan, Spain and the United States. Table C.1 shows the tested classification techniques, their accuracy and training time in seconds. We observed that the Decision Tree classifier (selected for the country and cultural dimension predictive modeling in chapter 6) enabled the highest accuracy with the smallest training time.

Classification Technique	Classification Rate	Training Time (seconds)
Decision Tree Classifier	100	0,003
Extra Trees Classifier	100	0,032
Multinomial NB	39,75	0,001
Nearest Centroid	28,61	0,002
Random Forest Classifier	98,76	0,026

Table C.1: Preliminary tests for selecting country-predictive classification model

C.2 Confusion Matrices for Country and Culture Group Predictions

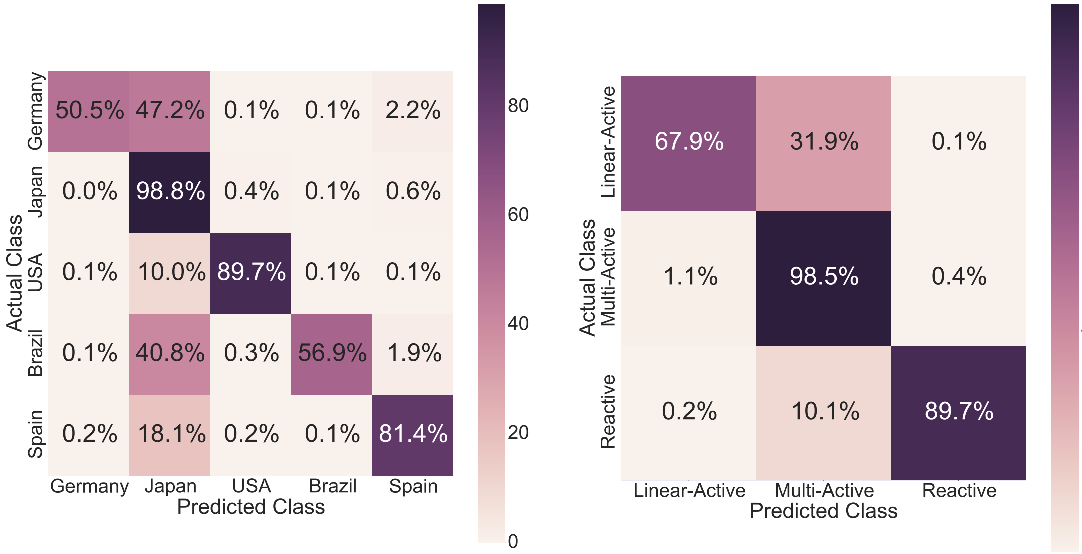
In this section, we perform an error analysis using confusion matrices for country and culture-group prediction experiments outlined in chapter 6 “User Origin Prediction”, performance results in Table 6.3. All tests were realised with Decision Trees in MATLAB. The confusion matrices were drawn with the Seaborn function heatmapseaborn.heatmap.

Following confusion matrices depict for each actual class break down of the predicted classes (in percent) when using LANG feature (language defined in the user profile) in Figure C.1, DEF (behaviour feature set using statistics of the Twitter feature usage) in Figure C.2 and the feature set combining both LANG + DEF in Figure C.3. It is important to mention that we had comparable group sizes, with the largest user group from Spain (26%), and smaller user group from the USA (11%) as shown in Figure 6.1. Therefore, our classification trees performed better than the classifier predicting the most populated class in all the cases.

As we see from Figure C.1, LANG feature set usage results in quite a weak prediction performance for “Germany” and “Brazil” classes. This could be explained by the large fraction of English language defined in Twitter user profiles for persons microblogging from Brazil and Spain (Figure 6.4 (a)). Figures C.2 and C.3 show more robust prediction results showing smaller rates of misclassified predictions for all classes.

C.3 Microsoft Azure

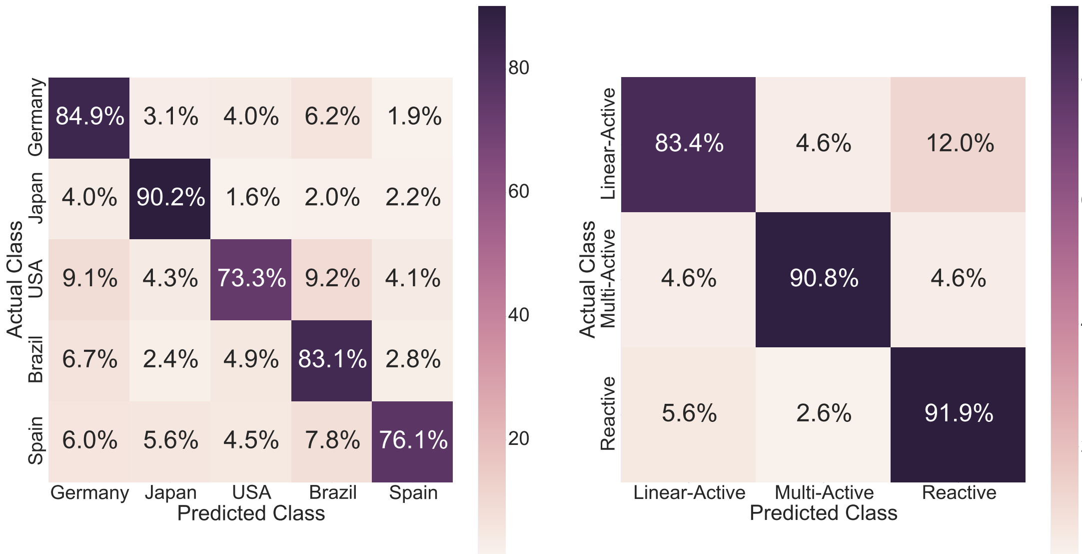
In this section, we briefly outline the country classification tests we performed on BEHAVIOUR feature set using Microsoft Azure Machine learning platform provided as a gift for this research project. This platform did not require any programming effort and was used for confirming the feasibility of country prediction outlined in [121] with the social data gathered from Twitter microblogs.



(a) Country Groups

(b) Cultural Groups

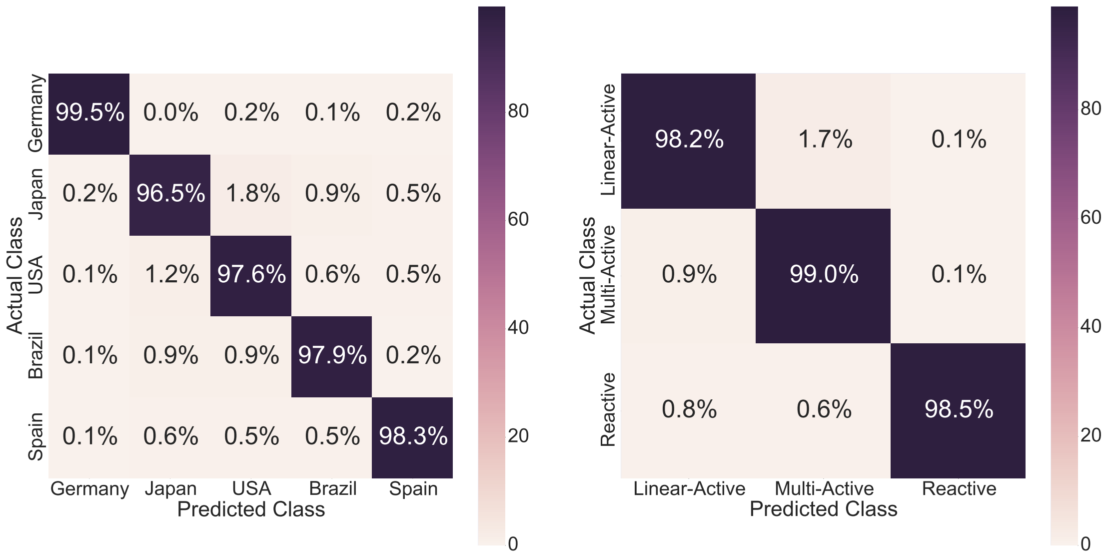
Figure C.1: Country and cultural dimension prediction confusion matrices: tests 1 and 2 (using language defined in the user profile)



(a) Country Groups

(b) Cultural Groups

Figure C.2: Country and cultural dimension prediction confusion matrices: tests 3 and 4 (using DEF feature set based on the behaviour Twitter feature)



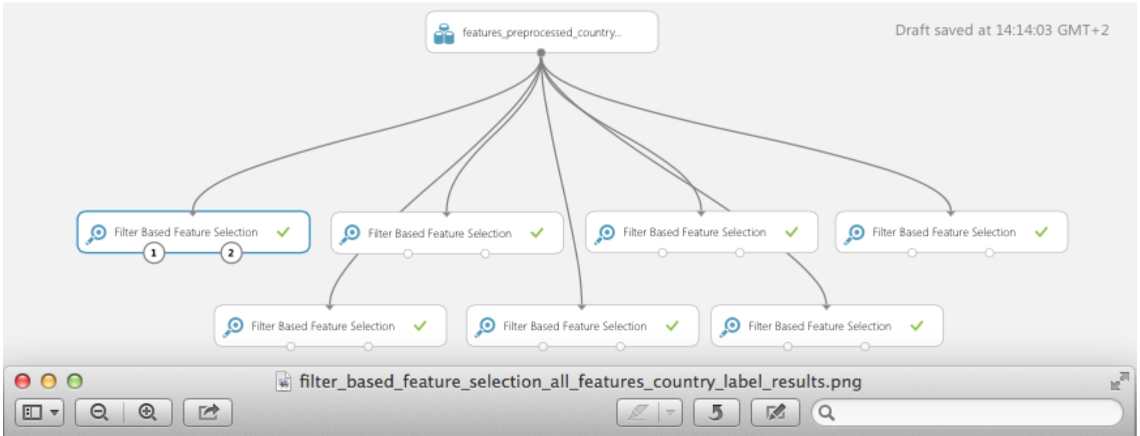
(a) Country Groups

(b) Cultural Groups

Figure C.3: Country and cultural dimension prediction confusion matrices: tests 5 and 6 (using LANG, which is language defined in the user profile and DEF feature set based on the behaviour Twitter feature)

C.3.1 Features

Figure C.5 shows the distribution of features analysed in [121]. While performing features selection in accord with the recommendations in [219] and with help of Azure Filter-based feature selection experiment, we found the features' importance depicted in Figure C.4. We notice that for the majority of performed methods, language, languages, tags, mentions, retweets, URLs are in the top features list, while features such as replies, friends, and followers mostly the least important features. The Filter-based feature selection is, however, might be not necessary for the experimental setting we employ. The decision tree models allow us to find out the most important features located at the top branches of the tree. As explained in [89], feature selection is not required as a pre-processing step when using decision trees already automatically defining the most important features.



	language	languages	tags	retweets	mentions	mobility	urls	weekends	replies	friends	followers
Chi Squared	29984	9397	4653	2589	2353	1266	700	535	512	467	375
	languages	mentions	retweets	tags	mobility	weekends	urls	replies	followers	friends	language
Pearson Correlation	0.6863	0.393401	0.364051	0.360429	0.273048	0.196676	0.13382	0.122435	0.110615	0.080838	0
	language	languages	tags	retweets	mentions	mobility	urls	weekends	replies	friends	followers
Mutual Information	0.963545	0.333742	0.180496	0.098689	0.094832	0.044094	0.026493	0.021175	0.02007	0.017403	0.013943
	tags	urls	languages	mobility	weekends	friends	followers	mentions	replies	retweets	
Count based	11778	12973	9989	10250	13084	13112	13111	13046	12708	11850	
	Languages	mentions	retweets	tags	mobility	weekends	urls	replies	followers	friends	language
Fisher Score	0.890385	0.183102	0.152782	0.149305	0.080561	0.040238	0.018234	0.015218	0.012387	0.006578	0
	languages	tags	mentions	retweets	mobility	weekends	followers	replies	urls	friends	language
Spearman Correlation	0.608244	0.527739	0.416567	0.404018	0.249819	0.174874	0.12151	0.117707	0.10472	0.049483	0
	languages	tags	mentions	retweets	mobility	weekends	followers	replies	urls	friends	language
Kendall Correlation	0.520364	0.408222	0.309927	0.308174	0.222647	0.130293	0.090421	0.087089	0.075949	0.035789	0

Figure C.4: Feature Importance

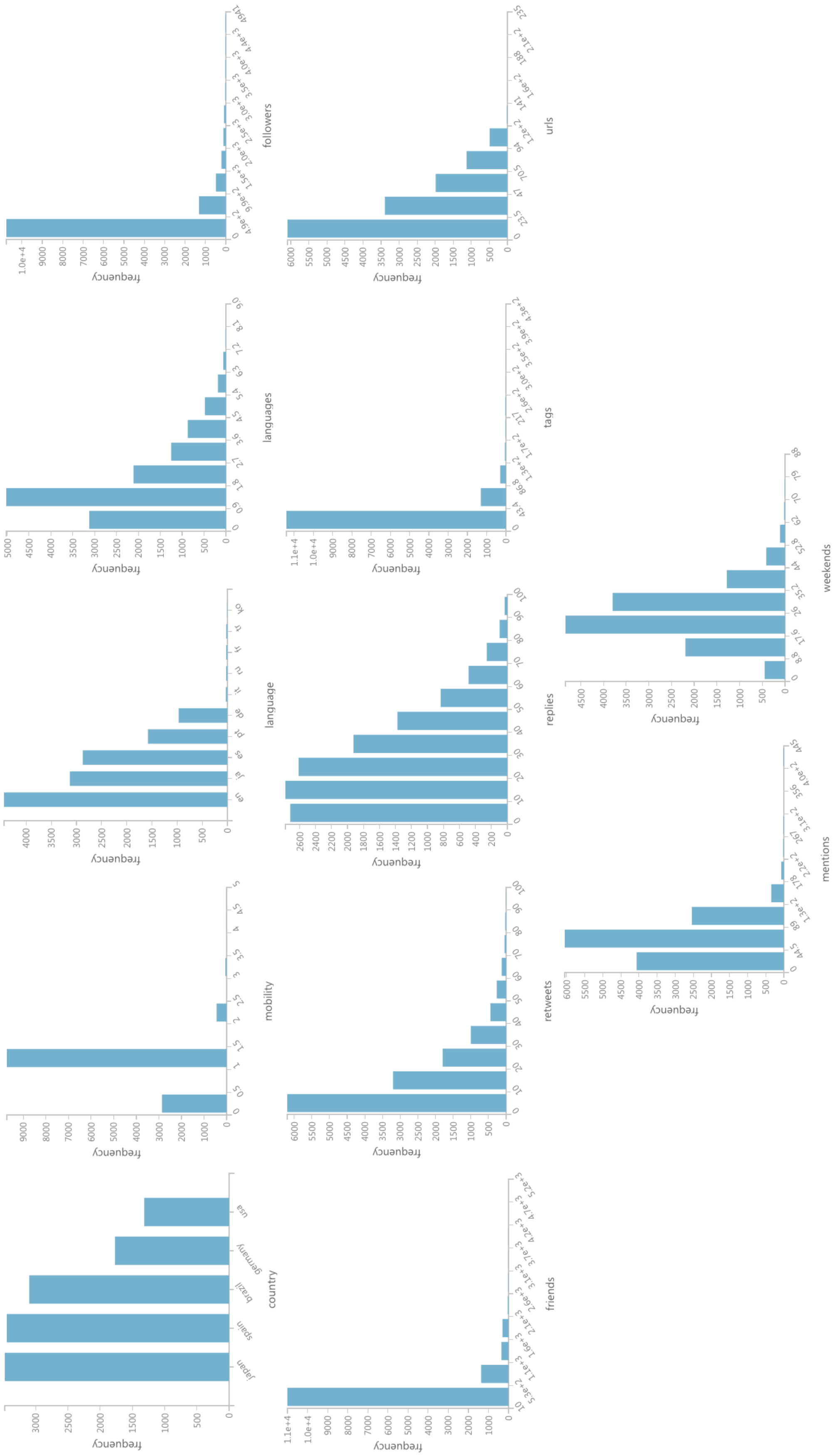


Figure C.5: Frequency Distributions of All Features (LANG+DEF feature set described in chapter 6)

C.3.2 Creating Classifiers in MS Azure

In this section we describe machine learning experiments performed in Microsoft Azure Studio and using “Features” dataset (see Table D.2). The test and train dataset split is 10% and 90% respectively. The label variables included country (’germany’, ’usa’, ’brazil’, ’spain’ and ’japan’) and culture groups (’LA’, ’MA’ and ’RE’). We compared Support Vector Machine (SVM) and Decision Forest classification models using all features, including language code and DEF as explained in [121] and chapter 6. Figure C.6 shows the experimental graph for SVM and Decision Forest multiclass classifiers. The classification modules’ settings are shown in Figure C.9. Confusion matrices for both models are shown in Figures C.14.

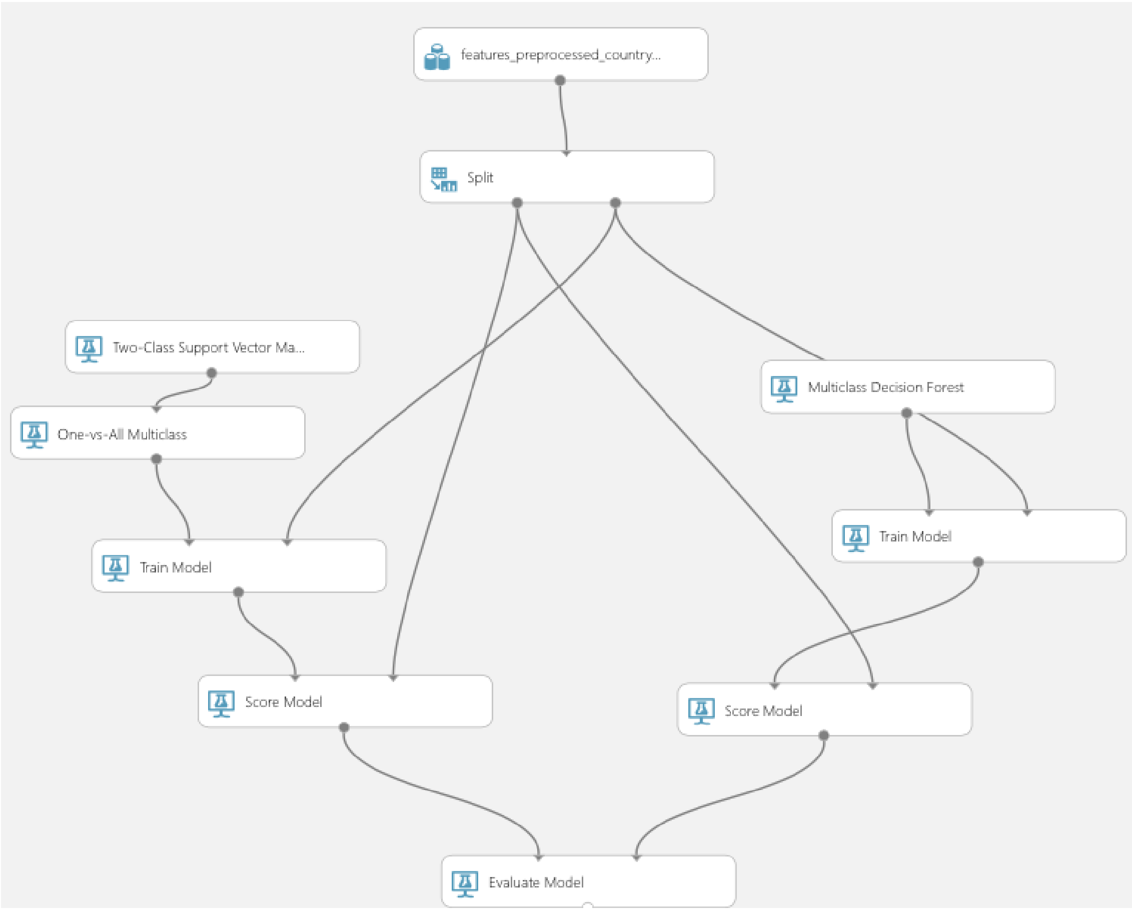


Figure C.6: Multiclass SVM and decision forest (country and culture groups prediction)

Setting	Value
Num Iterations	1
Lambda	0.001
Normalize Features	True
Perform Projection	False
Allow Unknown Levels	True
Random Number Seed	

Figure C.7: Two-class SVM and one-vs-all multiclass (untrained)

Setting	Value
Min Leaf Sample Count	1
Random Split Count	128
Max Depth	32
Ensemble Element Count	8
Class Count	5
Resampling Method	Bagging
Random Number Seed	5
Allow Unknown Levels	True

Figure C.8: Multiclass gemini decision forest classifier (untrained)

Figure C.9: Classification modules' settings

Predicted Class

Actual Class	brazil	germany	japan	spain	usa
brazil	94.8%	1.0%		1.6%	2.6%
germany	20.3%	65.5%		1.7%	12.4%
japan	6.9%	0.6%	90.6%		1.9%
spain	16.0%	2.3%		78.9%	2.9%
usa	12.9%	0.9%	3.4%	0.9%	81.9%

Figure C.10: Confusion matrix for multiclass SVM

Overall accuracy	0.842868
Average accuracy	0.937147
Micro-averaged precision	0.842868
Macro-averaged precision	0.84098
Micro-averaged recall	0.842868
Macro-averaged recall	0.823339

Figure C.11: Performance of multiclass SVM

Predicted Class

Actual Class	brazil	germany	japan	spain	usa
brazil	88.6%	5.6%	0.3%	5.6%	
germany	19.2%	69.5%	2.3%	4.0%	5.1%
japan	4.1%	0.3%	92.3%	1.4%	1.9%
spain	6.9%	5.4%	0.3%	86.9%	0.6%
usa	6.9%	6.0%	3.4%	4.3%	79.3%

Figure C.12: Confusion matrix for decision forest

Overall accuracy	0.857361
Average accuracy	0.942944
Micro-averaged precision	0.857361
Macro-averaged precision	0.842623
Micro-averaged recall	0.857361
Macro-averaged recall	0.832973

Figure C.13: Performance of decision forest

Figure C.14: Country prediction performance results on all features (LANG+DEF)

Appendix D

Datasets

In this section, we describe datasets collected using Twitter API during the project. In Table D.1 we provide a summary of all datasets, related publications, and chapters in the thesis. The overall database size is around 32GB.

Dataset	Description	Related Chapters / Publications
Features	Aggregated user profiles for the selected users tweeted from Brazil, Japan, USA, Spain and Germany, collected in 2012	chapter 6 “User Origin Prediction” / [121, 119]
Communication	Communication patterns research for top countries in Twitter, collected in 2014	chapter 7 “Communication Preferences” / [52]
Privacy	Twitter privacy settings usage data for a half of year data collection performed in 2015	chapter 8 “Privacy Settings Usage in Twitter” / [53]
Recommender	Movie ratings collection and recommender prototype developed for offline performance evaluations, collected in 2016	chapter 9 “Culture-aware Social Recommenders” / [51]

Table D.1: Information on datasets

D.1 Features Dataset

Table D.2 shows descriptive statistics for "Features" dataset representing aggregated Twitter user profiles. Besides, we also retain the country of the first tweet origin for each of our user profiles (Table 6.1 shows the number of users per country group). The reasoning behind collecting the data and feature descriptions are explained on pages 90-94.

	Feature	Minimum	Maximum	Mean	Std. Deviation
1	Hashtags	0	434	20.81	28.57
2	URLs	0	235	33.52	27.10
3	Languages Detected	0	9	1.56	1.48
4	Mobility	0	5	0.83	0.51
5	Weekend	0	88	25.43	9.89
6	Friends	10	5227	332.71	419.35
7	Followers	0	4941	345.96	548.63
8	User Mentions	0	445	64.53	36.54
9	Retweets	0	100	15.17	14.79
10	Replies	0	100	26.16	19.06

Table D.2: Descriptive statistics for data table Features (11,998 records)

D.2 Communication Dataset

Table D.3 shows summary of the "Profiles" table in the "Communication" dataset. The data collection process is described on page 119. Besides the presented numerical features we stored user numbers, language defined in the user profile, most used time zone, user country and cultural dimension associated. Using messages of followers, we store the most frequent language, country and cultural dimension in the followers' network. Table D.4 shows descriptive statistics for features stored in the "Communication" data table.

	Minimum	Maximum	Mean	Std. Deviation
User-related features related to BEHAVIOUR feature set (described on page 120)				
Languages	0	32	3.22	2.90
Hashtags	0	2699	17.11	86.16
URLs	0	1000	8.39	30.97
Replies	1	327	11.77	16.30
Retweets	1	1	1.00	0.00
Timezones	0	2	0.73	0.44
User Mentions	0	6922	95.97	186.17
Weekends	0	990	64.26	95.82
Tweets published	1	1000	118.80	156.33
Countries	1	27	1.01	0.27
Followers	0	713432	894.27	7598.80
Friends	0	204802	589.97	2731.45
Follower-related features for FOLLOWERS features set (described on page 122)				
FTimezones	0	40	2.42	2.26
FCountries	0	21	0.41	0.62
FLanguages	0	18	1.38	1.27
FInfluence	0	1	0.50	0.17

Table D.3: Descriptive Statistics for data table Profiles (13,289 records)

	Feature	Minimum	Maximum	Mean	Std. Deviation
1	Mobility	0	100	1.01	2.71
2	Languages	0	100	3.19	3.61
3	Tweets	1	1000	278.67	274.50
4	URLs	0	191	8.46	18.78
5	Retweets	0	100	1.14	3.01
6	Replies	0	100	9.36	8.14
7	Weekends	0	100	56.51	17.47
8	Followers	2	1224	351.17	196.32
9	Friends	0	1860	274.90	170.73
10	Hashtags	0	700	23.24	79.16
11	User Mentions	0	380	72.21	34.79

Table D.4: Descriptive statistics for data table Communication (107,960 records)

D.3 Privacy Dataset

Table D.5 shows a summary of the "Privacy Aggregated" table in the "Privacy" dataset. The data collection process is described on page 142. Additionally, we retain user geographic locations when available, inferred origins (based on the country location detection using "PLACE" features set) and privacy settings.

	Minimum	Maximum	Mean	Std. Deviation
FOLLOWERS	0	113746	350.12	1755.56
FRIENDS	0	81965	400.10	1457.69
INFLUENCE	0	5578.5	1.63	41.13
STATUSES	0	378168	1840.78	7307.81
LISTED	0	437	1.33	7.15
FAVOURITES	0	101704	454.39	2150.78
CHANGES	1	4	1.22	0.51
SOURCES	0	5	1.04	0.21

Table D.5: Descriptive statistics for "Privacy Aggregated" table (based on 21,133 user profiles)

D.4 Recommender Dataset

Table D.6 shows a summary of the "ALL" table in the "Recommender" dataset comprising all movie ratings collected. Additionally, we store user and movie identification numbers, user geographic locations when available, timezone, language defined in the user profile, inferred origins (based on the country location detection using "PLACE" features set), IMDB user votes, movie genre and year of a movie release. Additionally, Table J.1 (below) shows counts of movie ratings, users published these ratings, movies tweeted, countries identified and languages defined in the user profiles of tweet authors.

	Minimum	Maximum	Mean	Std. Deviation
RATING	0	10	7.63	1.85
LISTED	0	3303	19.24	111.10
FOLLOWERS	0	283805	1063.05	4582.35
FRIENDS	0	75159	563.03	1589.13

Table D.6: Descriptive statistics for "ALL" movie ratings table (39596 IMDB movie ratings)

Tables D.7 and D.8 summarise recommender performance results in timeline and offline user tests.

	Minimum	Maximum	Mean	Std. Deviation
NDCG	0.11	1.00	0.80	0.20
RMSE	1.35	1.94	1.65	0.14
R^2	-0.02	0.44	0.20	0.13
RATINGS #	3484	30716	18063.83	8246.65
MOVIES #	1802	6871	4686.36	1419.45
USERS #	1937.00	6702.00	4967.17	1352.52
VARIANCE	0	0.44	0.20	0.13

Table D.7: 'Timeline" tests table summary (350 rows)

	Minimum	Maximum	Mean	Std. Deviation
NDCG	0.14	1.00	0.91	0.12
RMSE	0	6.79	1.11	0.64
R^2	-64.71	1.00	-0.35	2.04
RATINGS #	645.00	2809.00	1868.65	937.62
VARIANCE	-28.32	1.00	0.26	0.81

Table D.8: "Offline user tests" summary (22,942 rows)

Appendix E

Geography and Cultural Dimensions

The following table shows the main geographic regions and their respective cultural dimensions in accord with the Lewis Model of Cultures. The cultural dimensions were manually assigned by human assessors as follows: Linear-Active (LA), Multi-Active (MA), Reactive (RE), Undefined (UN).

To assess the reliability of human raters agreement, we computed Krippendorff's α coefficient with the help of Python package developed by Thomas Grill [91]. Krippendorff's α was selected since it can deal with the missing data as described in [140](we estimated that almost 3 percent fo our dataset misses annotations). Krippendorff's α coefficient value of 0.91 was indicating good reliability following Krippendorff's agreement level of at least .80 threshold for good reliability α values [140].

#	Language name	Native name	Lang. ISO 639-1 code	1st Asses-sor	2nd Asses-sor	Place	Agreeed on
1	Northwest Caucasian	Abkhaz	ab	MA	MA	Russia, Slovakia	MA
2	Afro-Asiatic	Afar	aa	MA			LA
3	Indo-European	Afrikaans	af	MA			LA
4	Niger Congo	Akan	ak	MA	MA	Sub-Sahara African	MA
5	Indo-European	Albanian	sq	MA	MA	Russia / Slovakia	MA
6	Afro-Asiatic	Amharic	am	MA			LA
7	Afro-Asiatic	Arabic	ar	MA	MA	Saudi Arabia	MA
8	Indo-European	Aragonese	an	MA			LA
9	Indo-European	Armenian	hy	MA	MA		MA
10	Indo-European	Assamese	as	MA			LA
11	Northeast Caucasian	Avaric	av	MA			LA
12	Indo-European	Avestan	ae	MA			LA
13	Aymaran	Aymara	ay	MA			LA
14	Turkic	Azerbaijani	az	MA	MA	Turkey	MA
15	Niger Congo	Bambara	bm	MA	MA	Sub-Sahara African	MA

Table E.1: continued on the next page

#	Language name	Native name	Lang. ISO 639-1 code	1st Assessor	2nd Assessor	Place	Agreeed on
16	Turkic	Bashkir	ba	MA	MA	Turkey	MA
17	Language isolate	Basque	eu	MA	MA	Spain France	MA
18	Indo-European	Belarusian	be	MA	MA	Russia / Slovakia	MA
19	Indo-European	Bengali; Bangla	bn	MA			LA
20	Indo-European	Bihari	bh	MA			LA
21	Creole	Bislama	bi	MA			LA
22	Indo-European	Bosnian	bs	MA	MA	Russia / Slovakia	MA
23	Indo-European	Breton	br	MA	MA	France	MA
24	Indo-European	Bulgarian	bg	MA	MA	Russia / Slovakia	MA
25	Sino-Tibetan	Burmese	my	RE	RE	Thai China	RE
26	Indo-European	Catalan;Valencian	ca	MA	MA	Spain	MA
27	Austronesian	Chamorro	ch	MA			LA
28	Northeast Caucasian	Chechen	ce	MA	MA	Russia / Slovakia	MA
29	Niger Congo	Chichewa; Chewa; Nyanja	ny	MA	MA	Sub-Sahara African	MA
30	Sino-Tibetan	Chinese	zh	RE	RE	China	RE
31	Turkic	Chuvash	cv	MA	MA	Turkey	MA

Table E.1: continued on the next page

#	Language name	Native name	Lang. ISO 639-1 code	1st Assessor	2nd Assessor	Place	Agreeed on
32	Indo-European	Cornish	kw	LA	LA	UK	LA
33	Indo-European	Corsican	co	MA	MA	France Italy	MA
34	Algonquian	Cree	cr	MA			LA
35	Indo-European	Croatian	hr	MA	MA	Slovakia	MA
36	Indo-European	Czech	cs	LA	LA		LA
37	Indo-European	Danish	da	LA	LA		LA
38	Indo-European	Divehi; Dhivehi; Maldivian;	dv	MA			LA
39	Indo-European	Dutch	nl	LA	LA		LA
40	Sino-Tibetan	Dzongkha	dz	RE	RE	China	RE
41	Indo-European	English	en	LA	LA		LA
42	Constructed	Esperanto	eo	UN			LA
43	Uralic	Estonian	et	LA	LA	Finland	LA
44	Niger Congo	Ewe	ee	MA	MA	Sub-Sahara African	MA
45	Indo-European	Faroese	fo	LA	LA	Denmark	LA
46	Austronesian	Fijian	fj	RE			LA
47	Uralic	Finnish	fi	LA	LA		LA

Table E.1: continued on the next page

#	Language name	Native name	Lang. ISO 639-1 code	1st Asses-sor	2nd Asses-sor	Place	Agreeed on
48	Indo-European	French	fr	MA	MA		MA
49	Niger Congo	Fula; Fulah; Pulaar; Pular	ff	MA	MA	Sub-Sahara African	MA
50	Indo-European	Galician	gl	MA	MA	Spain	MA
51	South Caucasian	Georgian	ka	MA	MA	Russia / Slovakia	MA
52	Indo-European	German	de	LA	LA		LA
53	Indo-European	Greek, Modern	el	MA	MA	Turkey Italy	MA
54	Tupian	Guaran	gn	MA			LA
55	Indo-European	Gujarati	gu	MA			LA
56	Creole	Haitian; Haitian Creole	ht	MA	MA	France	MA
57	Afro-Asiatic	Hausa	ha	MA			LA
58	Afro-Asiatic	Hebrew(modern)	he	MA	MA	NOT SURE	MA
59	Niger Congo	Herero	hz	MA	MA	Sub-Sahara African	MA
60	Indo-European	Hindi	hi	MA	MA	India	MA
61	Austronesian	Hiri Motu	ho	RE			LA
62	Uralic	Hungarian	hu	MA	LA	Finland	LA
63	Constructed	Interlingua	ia	UN			LA

Table E.1: continued on the next page

#	Language name	Native name	Lang. ISO 639-1 code	1st Assessor	2nd Assessor	Place	Agreeed on
64	Austronesian	Indonesian	id	RE	RE		RE
65	Constructed	Interlingue	ie	UN			LA
66	Indo-European	Irish	ga	LA	LA		LA
67	Niger Congo	Igbo	ig	MA	MA	Sub-Sahara African	MA
68	Eskimo Aleut	Inupiaq	ik	RE			LA
69	Constructed	Ido	io	UN			LA
70	Indo-European	Icelandic	is	LA	LA	Denmark	LA
71	Indo-European	Italian	it	MA	MA		MA
72	Eskimo Aleut	Inuktitut	iu	RE			LA
73	Japonic	Japanese	ja	RE	RE		RE
74	Austronesian	Javanese	jv	MA			LA
75	Eskimo-Aleut	Kalaallisut Greenlandic	kl	RE			LA
76	Dravidian	Kannada	kn	MA			LA
77	Nilo-Saharan	Kanuri	kr	MA			LA
78	Indo-European	Kashmiri	ks	MA			LA
79	Turkic	Kazakh	kk	RE	MA	Turkey	MA

Table E.1: continued on the next page

#	Language name	Native name	Lang. ISO 639-1 code	1st Assessor	2nd Assessor	Place	Agreeed on
80	Austroasiatic	Khmer	km	RE	RE	Thai	RE
81	Niger Congo	Kikuyu Gikuyu	ki	MA	MA	Sub-Sahara African	MA
82	Niger Congo	Kinyarwanda	rw	MA	MA	Sub-Sahara African	MA
83	Turkic	Kyrgyz	ky	MA	MA	Turkey	MA
84	Uralic	Komi	kv	MA	LA	Finland	LA
85	Niger Congo	Kongo	kg	MA	MA	Sub-Sahara African	MA
86	Language isolate	Korean	ko	RE	RE		RE
87	Indo-European	Kurdish	ku	MA	MA	Turkey-Iran	MA
88	Niger Congo	Kwanyama Kuanyama	kj	MA	MA	Sub-Sahara African	MA
89	Indo-European	Latin	la	MA	MA	Italy	MA
90	Indo-European	Luxembourgish, Letzeburgesch	lb	LA	MA	France Belgium	LA
91	Niger Congo	Ganda	lg	MA	MA	Sub-Sahara African	MA
92	Indo-European	Limburgish, Limburgan, Limburger	li	LA	LA	Netherlands	LA
93	Niger Congo	Lingala	ln	MA	MA	Sub-Sahara African	MA
94	Tai Kadai	Lao	lo	MA	RE		LA

Table E.1: continued on the next page

#	Language name	Native name	Lang. ISO 639-1 code	1st Asses-sor	2nd Asses-sor	Place	Agreeed on
95	Indo-European	Lithuanian	lt	MA	MA	Russia / Slovakia	MA
96	Niger-Congo	Luba-Katanga	lu	MA	MA	Sub-Sahara African	MA
97	Indo-European	Latvian	lv	MA	MA	Russia / Slovakia	MA
98	Indo-European	Manx	gv	MA			LA
99	Indo-European	Macedonian	mk	MA			LA
100	Austronesian	Malagasy	mg	RE			LA
101	Austronesian	Malay	ms	RE	RE		RE
102	Dravidian	Malayalam	ml	RE			LA
103	Afro-Asiatic	Maltese	mt	RE			LA
104	Austronesian	Mori	mi	RE			LA
105	Indo-European	Marathi	mr	MA			LA
106	Austronesian	Marshallese	mh	RE			LA
107	Mongolic	Mongolian	mn	RE	RE		RE
108	Austronesian	Nauru	na	RE			LA
109	Den Yeniseian	Navajo Navaho	nv	RE			LA
110	Indo-European	Norwegian Bokml	nb	LA	LA		LA

Table E.1: continued on the next page

#	Language name	Native name	Lang. ISO 639-1 code	1st Assessor	2nd Assessor	Place	Agreeed on
111	Niger Congo	North Ndebele	nd	MA	MA	Sub-Sahara African	MA
112	Indo-European	Nepali	ne	MA			LA
113	Niger Congo	Ndonga	ng	MA	MA	Sub-Sahara African	MA
114	Indo-European	Norwegian Nynorsk	nn	LA	LA		LA
115	Indo-European	Norwegian	no	LA	LA		LA
116	Sino-Tibetan	Nuosu	ii	RE	RE		RE
117	Niger Congo	South Ndebele	nr	MA	MA	Sub-Sahara African	MA
118	Indo-European	Occitan	oc	MA			LA
119	Algonquian	Ojibwe Ojibwa	oj	MA			LA
120	Indo-European	Old Church Slavonic, Church Slavic, Church Slavonic, Old Bulgarian, Old Slavonic	cu	MA			LA
121	Afro-Asiatic	Oromo	om	MA			LA
122	India-Asia	Oriya	or	MA			LA
123	Indo-European	Ossetian, Ossetic	os	MA	MA	Russia / Slovakia	MA
124	Indo-European	Panjabi, Punjabi	pa	MA			LA

Table E.1: continued on the next page

#	Language name	Native name	Lang. ISO 639-1 code	1st Asses-sor	2nd Asses-sor	Place	Agreeed on
125	Indo-European	Pli	pi	MA			LA
126	Indo-European	Persian(Farsi)	fa	MA			LA
127	Indo-European	Polish	pl	MA	MA		MA
128	Indo-European	Pashto, Pushto	ps	MA			LA
129	Indo-European	Portuguese	pt	MA	MA		MA
130	Quechuan	Quechua	qu	MA			LA
131	Indo-European	Romansh	rm	MA			LA
132	Niger Congo	Kirundi	rn	MA			LA
133	Indo-European	Romanian	ro	MA	MA	Russia / Slovakia	MA
134	Indo-European	Russian	ru	MA	MA		MA
135	Indo-European	Sanskrit	sa	MA			LA
136	Indo-European	Sardinian	sc	MA			LA
137	Indo-European	Sindhi	sd	MA			LA
138	Uralic	Northern Sami	se	MA			LA
139	Austronesian	Samoan	sm	MA			LA
140	Creole	Sango	sg	MA			LA

Table E.1: continued on the next page

#	Language name	Native name	Lang. ISO 639-1 code	1st Assessor	2nd Assessor	Place	Agreeed on
141	Indo-European	Serbian	sr	MA	MA	Russia / Slovakia	MA
142	Indo-European	Scottish Gaelic; Gaelic	gd	LA	LA		LA
143	Niger Congo	Shona	sn	MA	MA	Sub-Sahara African	MA
144	Indo-European	Sinhala, Sinhalese	si	MA			LA
145	Indo-European	Slovak	sk	MA	MA		MA
146	Indo-European	Slovene	sl	MA	MA	Russia / Slovakia	MA
147	Afro-Asiatic	Somali	so	MA	MA	Sub-Sahara African	MA
148	Niger Congo	Southern Sotho	st	MA	MA	Sub-Sahara African	MA
149	Turkic	South Azerbaijani	az	MA	MA	Turkey	MA
150	Indo-European	Spanish; Castilian	es	MA	MA		MA
151	Austronesian	Sundanese	su	MA			LA
152	Niger Congo	Swahili	sw	MA	MA	Sub-Sahara African	MA
153	Niger Congo	Swati	ss	MA	MA	Sub-Sahara African	MA
154	Indo-European	Swedish	sv	LA	LA		LA
155	Dravidian	Tamil	ta	MA		?	LA
156	Dravidian	Telugu	te	MA			LA

Table E.1: continued on the next page

#	Language name	Native name	Lang. ISO 639-1 code	1st Assessor	2nd Assessor	Place	Agreeed on
157	Indo-European	Tajik	tg	MA			LA
158	Tai Kadai	Thai	th	RE	RE		RE
159	Afro-Asiatic	Tigrinya	ti	RE			LA
160	Sino-Tibetan	Tibetan Standard, Tibetan Central	bo	RE	RE		RE
161	Turkic	Turkmen	tk	MA	MA	Turkey	MA
162	Austronesian	Tagalog	tl	RE			LA
163	Niger Congo	Tswana	tn	MA	MA	Sub-Sahara African	MA
164	Austronesian	Tonga(Tonga Islands)	to	MA			LA
165	Turkic	Turkish	tr	MA	MA	Turkey	MA
166	Niger Congo	Tsonga	ts	MA	MA	Sub-Sahara African	MA
167	Turkic	Tatar	tt	MA	MA	Turkey	MA
168	Niger Congo	Twi	tw	MA	MA	Sub-Sahara African	MA
169	Austronesian	Tahitian	ty	MA			LA
170	Turkic	Uyghur, Uighur	ug	MA	MA	Turkey	MA
171	Indo-European	Ukrainian	uk	MA	MA	Russia / Slovakia	MA

Table E.1: continued on the next page

#	Language name	Native name	Lang. ISO 639-1 code	1st Asses-sor	2nd Asses-sor	Place	Agreeed on
172	Indo-European	Urdu	ur	MA			LA
173	Turkic	Uzbek	uz	MA	MA	Turkey	MA
174	Niger Congo	Venda	ve	MA	MA	Sub-Sahara African	MA
175	Austroasiatic	Vietnamese	vi	RE	RE		RE
176	Constructed	Volapk	vo	UN			LA
177	Indo-European	Walloon	wa	MA	MA	Belgium-France	MA
	Indo-European	Welsh	cy	LA	LA		LA
178							
179	Niger Congo	Wolof	wo	MA	MA	Sub-Sahara African	MA
180	Indo-European	Western Frisian	fy	LA	LA	Netherlands	LA
181	Niger Congo	Xhosa	xh	MA	MA	Sub-Sahara African	MA
182	Indo-European	Yiddish	yi	MA		?	LA
183	Niger Congo	Yoruba	yo	MA	MA	Sub-Sahara African	MA
184	Tai Kadai	Zhuang, Chuang	za	MA		?	LA
185	Niger Congo	Zulu	zu	MA	MA	Sub-Sahara African	MA

Table E.1: continued on the next page

#	Language name	Native name	Lang. ISO 639-1 code	1st Assessor	2nd Assessor	Place	Agreeed on

Table E.1: Regions, languages and cultural dimensions assigned

Appendix F

Communication Analysis Sample

Table F.1 summarises sample of 4250109 tweets published by 3198307 users and collected from 17th to 18th of March 2014. This sample table is collected for communication preferences analysis discussed in chapter 7. Please note that the geographic coordinates were only available for 2% of tweets.

Country (ISO 3166-1 alpha-2)	Number of Users	Country-specific Language ISO code	Having Country-related Languages in the User Profiles
US	28167	en	27640
BR	10474	pt	8008
ID	6998	id	3041
TR	5110	tr	4803
GB	4758	en	4658
JP	3056	ja	2883
ES	2627	es	2396
FR	2450	fr	2295
MY	2340	ms, but English is wide-spread	ms=13, en=2289, total=2302
MX	1669	es	1399

Table F.1: Top 10 countries amongst users with geo-tagged tweets (sample data collected for "Communication" dataset)

Appendix G

Privacy Sample Data

Table G.1 presents tweets sample created during 3 days of listing to Twitter stream from 26th to 29th of November 2014 (see chapter 8). We observed that only 0.5% of tweets included country information in metadata of recently registered users. In total, we retained tweets from 49 defined by Twitter countries.

Country Code	Tweets	Users	Country Code	Tweets	Users
Undefined	48020	22450	BR	51	41
US	43	35	TR	27	23
ID	17	13	JP	18	11
GB	10	8	CO	8	7
SA	6	6	DO	7	5
IQ	5	5	ES	5	5
AR	5	5	IN	5	5
MX	4	4	FR	4	4
TH	3	3	MY	3	3
IT	4	3	OM	3	2
RU	2	2	KR	3	2
PE	2	2	DE	3	2
CA	2	2	JO	2	2
VE	1	1	EG	1	1
NL	1	1	TZ	1	1
CL	1	1	BA	1	1
PK	2	1	AE	1	1
HK	1	1	PA	1	1
YE	1	1	NO	1	1
UA	1	1	DZ	1	1
QA	1	1	NZ	1	1
EC	1	1	NG	1	1
TW	1	1	MA	1	1
SE	1	1	RO	1	1
AZ	1	1	PH	1	1

Table G.1: Number of users and tweets by country

Appendix H

Microblogs Origin Predictive Models' Comparison

To train our classification models, we selected geographically-tagged tweets having country locations provided in the tweet metadata. During several days of data collection (from 28/03/16 to 11/04/16), more than million of tweets were stored, out of which 150 thousand tweets were randomly selected for countries having defined cultural group defined and from which we had at least 100 tweets in our dataset as seen in Figure H.1. The created dataset was divided into three parts: training set of 90 thousand tweets, cross-validation and test sets of 30 thousand tweets each. It is important to mention, that as in the previous experiments we assume that majority of users tweet from their countries of origin.

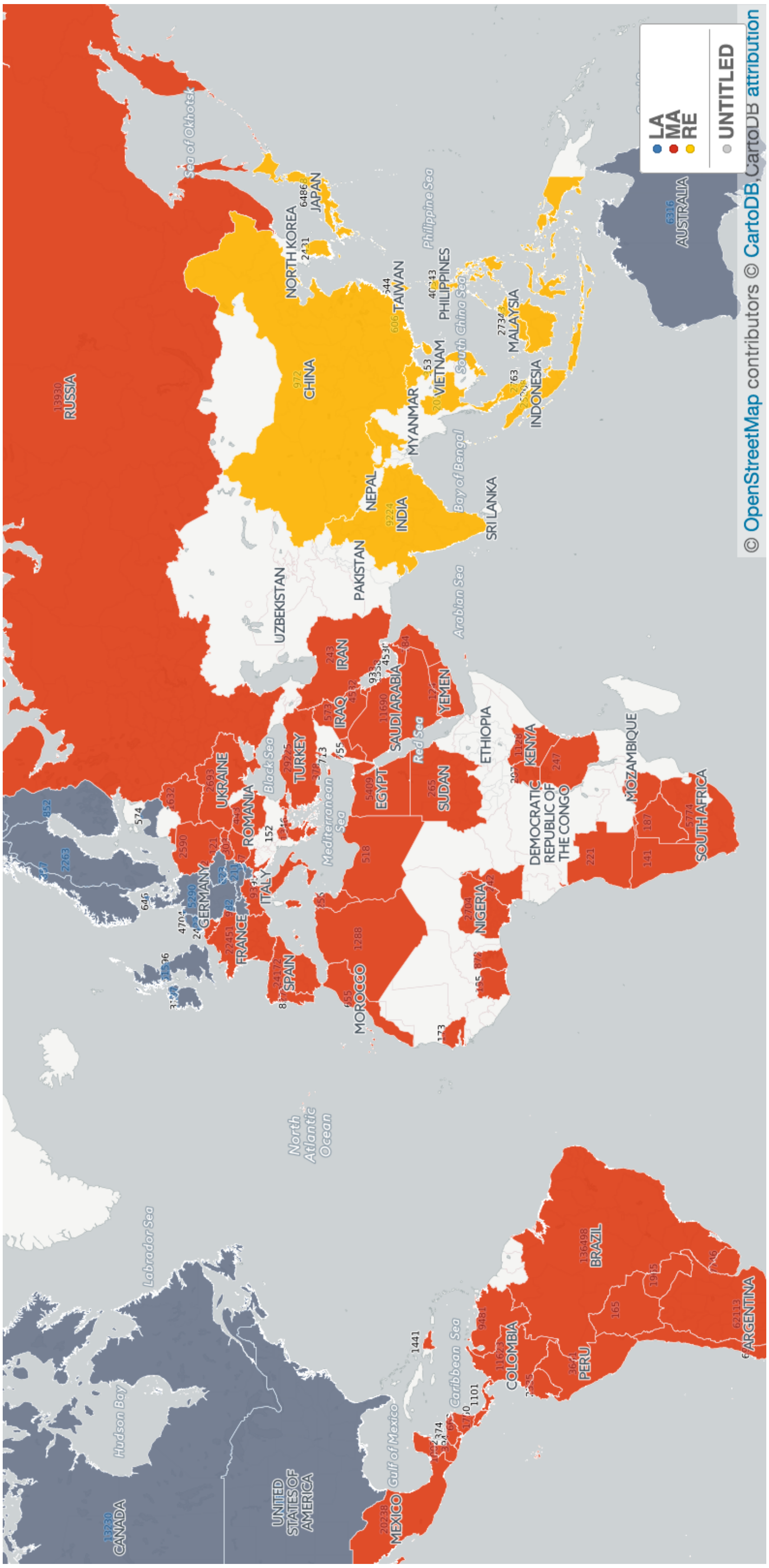


Figure H.1: Geography of origin predictions (map of countries)

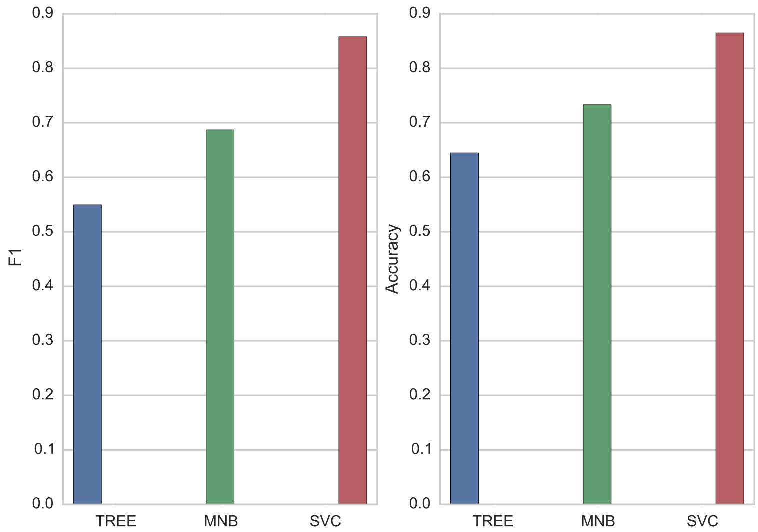


Figure H.2: Performance of the word-based baseline classification models

We analysed the performance of several classification approaches described in Table H. Overall, we exploited two feature sets: user language defined in the user profile ("LANGUAGE") and textual metadata including user language, time zone and free-text provided in user location ("PLACE" feature set).

Name	Description?
TREE	Decision Tree using user languages defined in Twitter profile for predicting user origins
SVC (WORD)	Support Vector Classification using PLACE feature set described above for creating word count vectors further transformed into Tfidf representations
SVC (CHAR)	Similarly to the SVC (WORD) using character n-grams
SVC (UNION)	SVC-based combining features of SVC (WORD) and SVC (CHAR)
MNB (WORD, CHAR, UNION)	The multinomial Naive Bayes created in three variations: based word n-grams (WORD), character n-grams (CHAR), and their combination (UNION).

Table H.1: Classification approaches

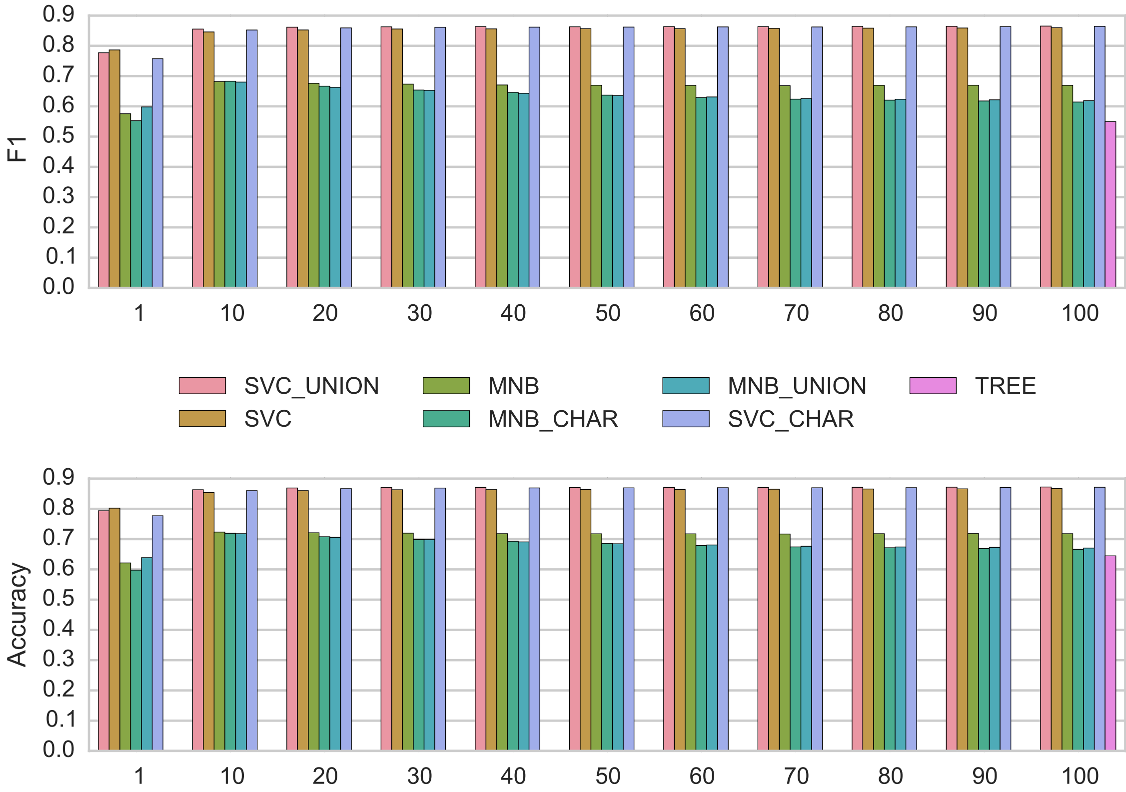


Figure H.3: Cross-Validation performance

Figure H.2 shows prediction accuracy and F1-measure average values for 10-times cross-validation performed for three baseline classification models trained for predicting 95 tweet origin countries. We observed a slight improvement of Accuracy and F1 metrics for SVC CHAR and UNION strategies over SVC for all top features percentiles we analysed except of 1st percentile in 10-times CV tests as seen in Figure H.3. In the additional test, we observed similar picture indicating also that usage of 10th percentile of all features allows comparable to the performance with all features extracted as show in Figure H.4. Nevertheless when tested on geo-enabled tweets with movie ratings, CHAR-based SVC did not show a superior Accuracy performance (Figure H.5), in further experiments we exploited CHAR SVC, which enabled lower test error than SVC WORD and smaller number of features as compared to SVC UNION (Figure H.6).

Since we have 95 countries in our predictive model, it was paramount to evaluate the performance of the particular classes. As we can see in Figure H.7(b) PLACE model enables better performance regarding sensitivity and specificity. Figure H.7(a)

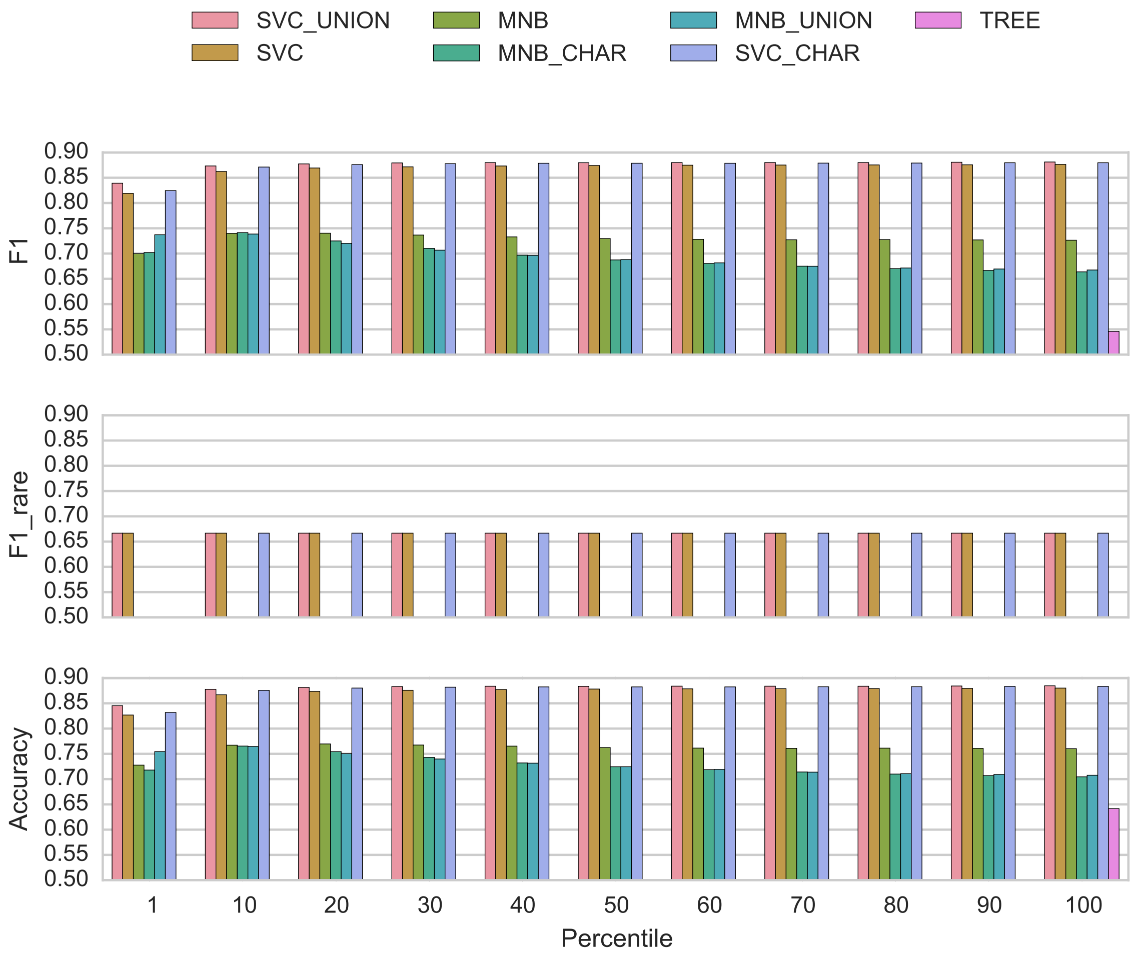


Figure H.4: Test Performance (F1_rare is the F1 metric for the rarest class having lowest performance values)

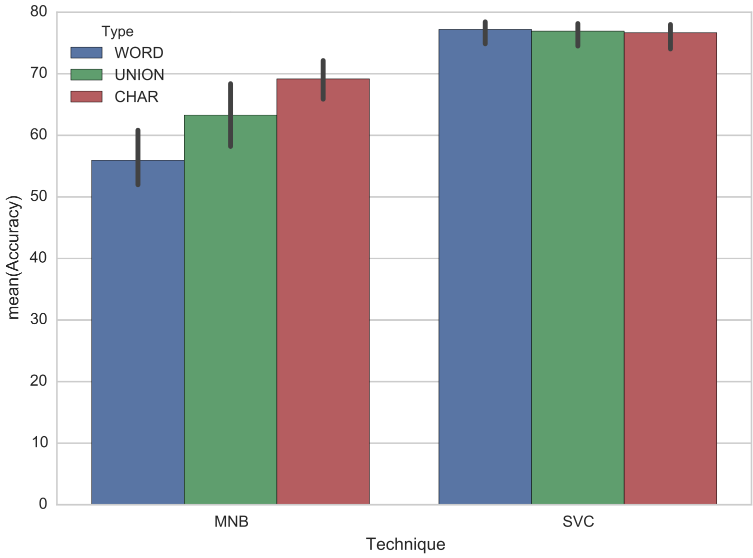


Figure H.5: Accuracy for geo-enabled tweets set

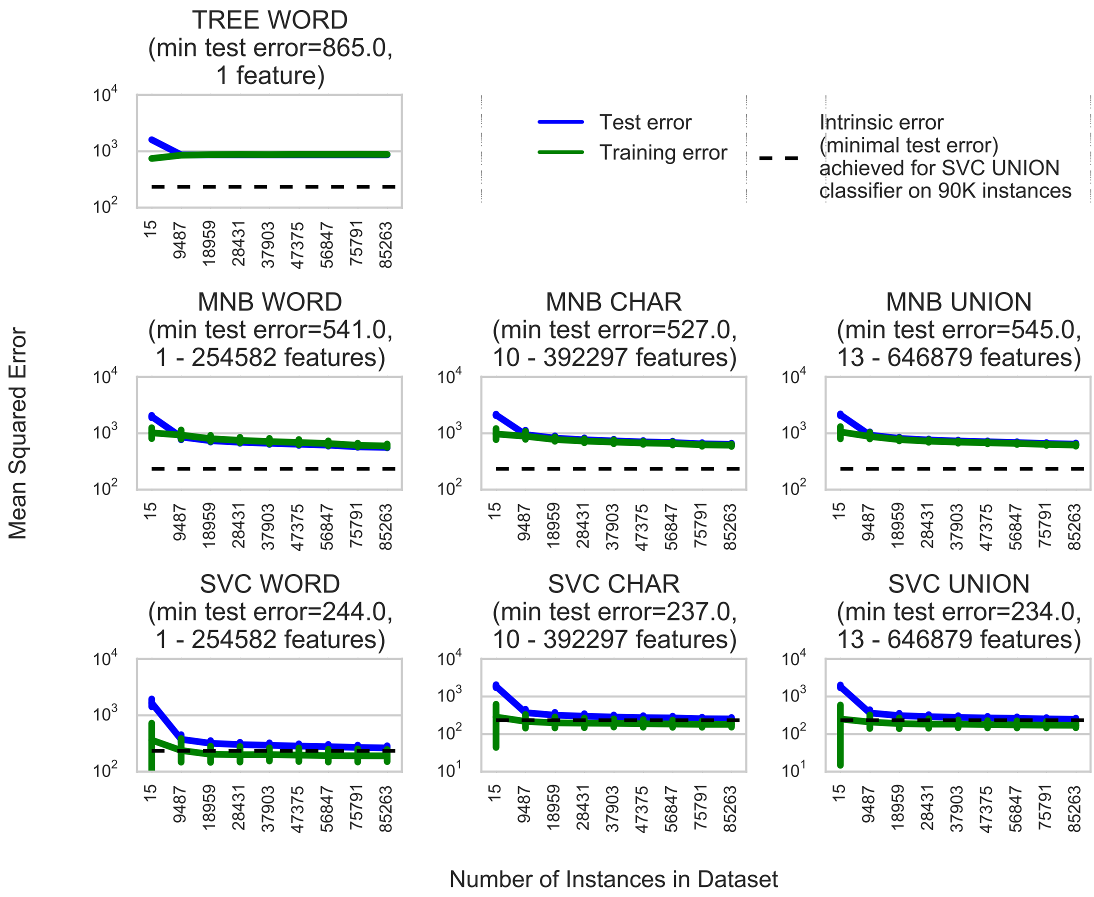
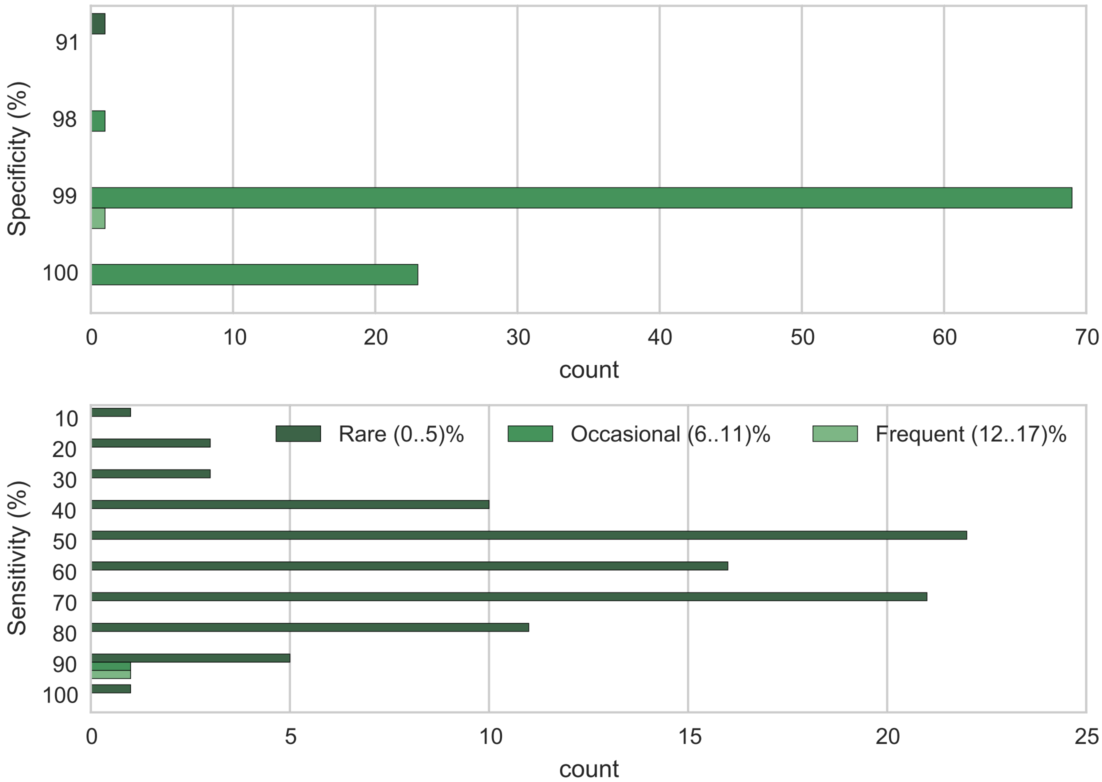


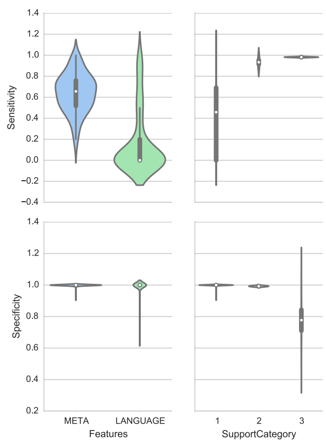
Figure H.6: Learning curves in test

shows that PLACE model enables the highest Sensitivity and Specificity for "Occasional" and "Frequent" classes, while "Rare" classes including 54 countries have Sensitivity of 60% and above.

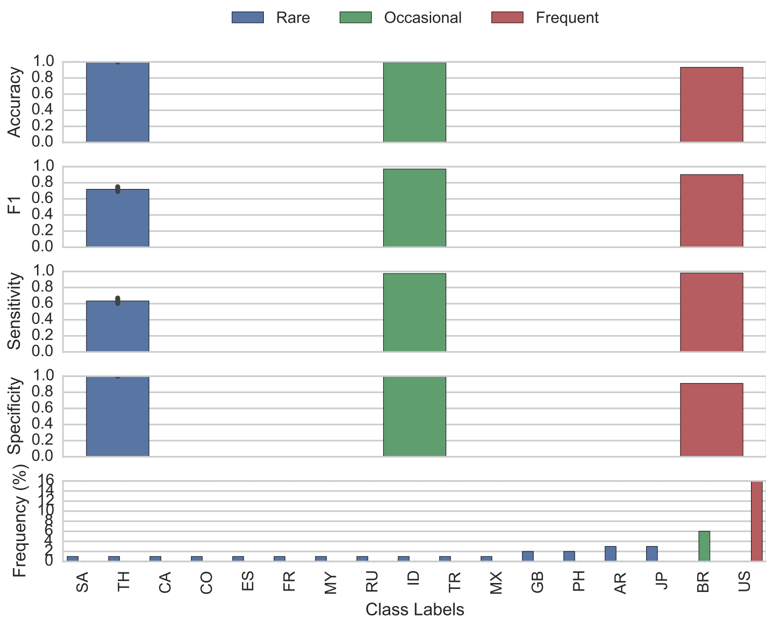
Overall, we predicted 27 "Rare" countries, which were included in less than 6% cases in the test dataset, with Sensitivity and Specificity of 75% and above. Figure H.7 shows average performance metrics summary for all frequency groups and top associated countries. In short, the countries predicted with the True Positive and Negative rates of 75% and above include Makedonia (MK), United States (US), Brazil (BR), Argentina (AR), Turkey (TR), Japan (JP), Russian Federation (RU), France (FR), Nicaragua (NI), Sweden (SE), Italy (IT), Senegal (SN), Costa Rica (CR), Indonesia (ID), Ghana (GH), Poland (PL), Saudi Arabia (SA), Venezuela (VE), Czech Republic (CZ), Greece (GR), Portugal (PT), Thailand (TH), Panama (PA), Spain (ES), Great Britain (GB), India (IN), Slovenia (SI), Kenya (KE) and Zimbabwe (ZW) in order of decreased True Positive Rate.



(a) Sensitivity and Specificity for Different Categories of Occurrence (SupportCategory above) in PLACE Model



(b) Sensitivity and Specificity for Different Models and Support Categories (1="Rare", 2="Occasional", 3="Frequent" as below)



(c) PLACE Model Performance and Top Occurred Class Labels

Figure H.7: Class Performance

H.1 Automatic country prediction performance

To create culture-aware recommendation strategies, we wanted to focus on a set of countries with good prediction performance. For this, we tested our data collection process for searching Twitter for a string "I rated IMDB" (similarly to the main data collection). We observed 29 countries with good prediction performance, out of which we selected 15 countries with user profiles and item data available with more than ten ratings in each. Their classification model support and country prediction performance is shown in Table H.2.

Country	Support	Sensitivity	Specificity	Accuracy	F1
Multi-Active					
Brazil (BR)	3863	0.97	1.00	0.99	0.97
Turkey (TR)	811	0.96	1.00	1.00	0.97
Russia(RU)	381	0.94	1.00	1.00	0.90
Italy (IT)	251	0.86	1.00	1.00	0.89
Saudi Arabia (SA)	331	0.81	1.00	0.99	0.77
Venezuela (VE)	266	0.81	1.00	1.00	0.87
Greece (GR)	27	0.78	1.00	1.00	0.88
Spain (ES)	695	0.76	1.00	0.99	0.82
Linear-Active					
US	9525	0.98	0.91	0.93	0.90
Sweden (SE)	63	0.87	1.00	1.00	0.87
Great Britain (GB)	1454	0.76	0.99	0.98	0.81
Reactive					
Japan (JP)	1863	0.94	1.00	0.99	0.96
Indonesia (ID)	673	0.83	1.00	1.00	0.88
Thailand (TH)	604	0.77	1.00	0.99	0.83
India (IN)	261	0.75	1.00	1.00	0.82

Table H.2: Selected countries' performance and cultural groups (support is the number of country occurrences in the test set)

H.2 Human assessment of tweet country origin predictions

Table H.3 shows country assignments for the particular tweets of movie ratings (accompanies by language defined in Twitter profile in "Language" column, the timezone in "Timezone" and free-text location field in "Location") by two human assessors (Rater1 and Rater2). Please note that the user country location provided by a user device (columns "Device") and automatically inferred country (column "Inferred") were initially hidden from the raters. The Timezone information is not always available and provided by the user device. Please consider that we removed user tweets and descriptions ("About me" field) from this table due to privacy concerns, however, the information was used by the human raters as openly available additional information solely for the country location assignments.

In chapter 5 on Methodology, we discussed and justified the selection of inter-rater agreement (or reliability) measures while referring to [42] for Cohen kappa, [74, 73] for Fleiss' kappa and Krippendorff's [140, 141] with their respective arbitrary benchmark scales. In Table H.4 (a) we show Krippendorff's (above diagonal) and agreement percentages (below diagonal), in Table (b) we show Fleiss' kappa (above diagonal) and Cohen kappa coefficients (below diagonal). We used several metrics to emphasise how the interpretation might be different in respect to selected metrics. For instance, as seen in Tables (a) and (b), Krippendorff's for Classifier and raters' assignments show lower IRR results, which are quite different from Fleiss and Cohen Kappa. We observe here not sufficiently high values for Krippendorff's (< 0.67), however, moderate good values for Fleiss' and Cohen's Kappa. Table (c) shows inter-rated reliability coefficients calculated for more than two annotators. This is useful for finding out the overall inter-rater agreement between all annotations.

Annotation results for two human raters, country location provided by Twitter ("Device" column) and the labels assigned by the country-predictive model ("Inferred" column)

Device	Language	Location	Timezone	Inferred	Rater1	Rater2	Rater 2, Comment
GB	en	Clacton-on-sea, Essex		GB	GB	GB	
US	en	Northern Cali		US	US	US	
GB	en	Stafford	London	GB	GB	GB	
US	ar	Where I meant to be.	Baghdad	US	IQ	IQ	
US	en	Sherman Oaks	Pacific Time (US & Canada)	US	US	US	
GB	en		London	GB	GB	GB	
US	en	41200.00%	Central Time (US & Canada)	US	US	CA	
RU	en	russia, barnaul	Novosibirsk	RU	RU	RU	
JP	en		Hawaii	US	JP	US	
ID	en	Pontianak , ID	Jakarta	ID	ID	ID	
RU	en	Ekaterinburg, Russia	Ekaterinburg	RU	RU	RU	
GB	en	Sheffield (Wednesday)	London	GB	GB	GB	
GR	en	Goudi	Athens	US	GR	GR	
US	en			US	X	X	ACCOUNT SUSPENDED
US	en	Hampton Roads, VA	Eastern Time (US & Canada)	US	US	US	
TH	en	13.873772,100.405297	Bangkok	US	TH	TH	
GB	en	Leeds, West Yorkshire	London	GB	GB	GB	
SA	en	Riyadh, Dhahran	Baghdad	US	IQ	IQ	
US	en			US	US	US	

Table H.3: Human assessment of the country predicting classifier

Device	Language	Location	Timezone	Inferred	Rater1	Rater2	Rater 2, Comment
US	en	Roseville, CA	Pacific Time (US & Canada)	US	US	US	
SA	en		Hawaii	US	US	US	
IN	en	New Delhi, India	New Delhi	US	IN	IN	
US	en	Chicago, ILL	Mountain Time (US & Canada)	US	US	US	
TR	en		Istanbul	TR	TR	TR	
GB	en	Huddersfield & Leeds, England	London	GB	GB	GB	
GB	en	Haywards Heath, England	London	GB	GB	GB	
US	en	Ponca City, Oklahoma	Central Time (US & Canada)	US	US	US	
US	en	Cyprus		US	CY	CY	
SA	en	SA, Sudayr	Baghdad	US	IQ	IQ	
US	en	Chicago	Central Time (US & Canada)	US	US	US	
TH	en	Phuket, Thailand		US	TH	TH	
GB	en	Sheffield, England		GB	GB	GB	
GB	en		Amsterdam	GB	GB	NL	
US	en	Nashville, TN	Central Time (US & Canada)	US	US	US	
GB	en	Plymouth, UK & Budapest, HU	London	GB	GB	GB	2ND LOC HU
US	en	New York	Eastern Time (US & Canada)	US	US	US	
GB	en	Pluto (Yes, it's a planet too)	West Central Africa	US	NG		WEST CENTRAL AFRICA ?
ES	en	Barcelona	Madrid	ES	ES	ES	

Table H.3: Human assessment of the country predicting classifier

Device	Language	Location	Timezone	Inferred	Rater1	Rater2	Rater 2, Comment
US	en	Vista, CA USA	Pacific Time (US & Canada)	US	US	US	
TR	en	world-wide	Istanbul	US	TR	TR	
GB	en	Tranent East Lothian Scotland	Edinburgh	GB	GB	GB	
SA	en	Jubail - Dhahran	Riyadh	US	SA	SA	
GR	en	Athens		US	GR	GR	
US	en			US	IN	IN	?RATES MOVIES FREQUENTLY
RU	ru	moscow, rf	Moscow	RU	RU	RU	
GB	en	iPhone: 53.203595,-0.606108	London	GB	GB	GB	
ES	en	Motril	Madrid	ES	ES	ES	
SA	ar	Camp Nou ❤ Jc	Baghdad	US	IQ	IQ	
US	en	San Antonio, TX 78217	Central Time (US & Canada)	US	US	US	
US	en			US	US	US	? US
US	en	cary, north carolina	Eastern Time (US & Canada)	X	US	US	
ID	id	Bogor, Indonesia	Jakarta	ID	ID	ID	
ID	en	Indonesia	Jakarta	ID	ID	ID	
US	en		Eastern Time (US & Canada)	US	US	US	? US
RU	en		Volgograd	RU	RU	RU	
SA	en	Al Khobar, Kingdom of Saudi Arabia		SA	SA	SA	

Table H.3: Human assessment of the country predicting classifier

a) Krippendorff's α (above diagonal) and agreement percentages (below diagonal)

	Rater 1	Rater 2	Classifier	Twitter
Rater 1		0.94	0.6	0.81
Rater 2	92.86%		0.59	0.76
Classifier	69.64%	67.86%		0.66
Twitter	83.93%	78.57%	75.00%	

To empasise almost excellent percentage results with values equal or above 90, we mark them in geen color, yellow and pink colors are used for showing good values in range [75, 90) and moderate values of at least 60% respectively. For inter-rater reliability coeffiecint alpha we show almost perfect values in green (at least 0.9 and above), good results (values in range [0.80, 0.9)) in yellow, moderate results (values in range [0.67, 80) in pink color, and poor results (values less that 0.67) were in white cells in accordance with the suggestion by Krippendorff [140].

b) Fleiss' kappa (above diagonal) and Cohen kappa coefficients (below diagonal)

	Rater 1	Rater 2	Classifier	Twitter
Rater 1		0.92	0.59	0.8
Rater 2	0.92		0.57	0.74
Classifier	0.61	0.58		0.65
Twitter	0.8	0.74	0.66	

To mark almost excellent kappa results with values equal or above .81, we mark them in geen color, yellow and pink colors are used for showing good values in range [.75, .81) and moderate values in range [.40, .75) respectively, referring to the benchmark scale by Landis and Koch in [145].

c) Inter-rater reliability coefficients for more than 2 annotators

Annotators	Fleiss' Kappa	Krippendorf's α
Human Raters + Classifier (Inferred)	0.7	0.72
Human Raters +Twitter (Device)	0.82	0.84
Human Raters +Classifier +Twitter (Device)	0.72	0.74

Table H.4: Summary of the human assessment of the country predicting classifier

(we mark almost perfect results of percentage and inter-rater reliability coefficeint values (as advised by Landis and Koch in [145] for Kappa values, and by Krippendorff [140] for α values while considering an additional level of α ≥.90 for excellent results) in green, good results in yellow and moderate results in pink color; in place of country names we show ISO 3166-1 alpha-2 codes)

Appendix I

Movie Genre Preferences in Respect to Locality Groups

Countries are inferred based on location, user language and time-zone in Twitter profiles; Lewis dimensions are associated with inferred countries using a dictionary. Significance level when applicable was denoted by ** with $p < 0.01$ and * with $p < 0.05$. To compare independent sample means, we employ t-test provided by Python Library Scipy [216]. This library allows performing t-test when variances are equal (by setting up parameter 'equal_var'=True), otherwise ('equal_var'=False)) calculating Welchs t-test. We assumed the sample variances equality with the help of Levene test [214]; we assumed that the variability is about the same when significance level $p > 0.05$. We also performed normality tests using normality test "based on D'Agostino and Pearson's test" [215] referring to [54] while considering $p > 0.05$ when the null hypothesis (that the sample comes from a normal distribution) cannot be rejected. Based on the normality test results, we observed that only in 41 out of 167 we could not reject the null hypothesis with $p > 0.05$ values, and we had 31 samples with too few test cases (less than 8 cases).

Test ID	Genre	Locality	Group1	N1	μ_1	σ_1	Group2	N2	μ_2	σ_1	t-stat	p-value	significance
38	Drama	country	SA	3472	8.07	1.93	other	6821	7.64	1.75	10.95	0.00	**
53	Drama	language	ar	2810	8.11	1.94	other	7483	7.66	1.76	10.73	0.00	**
74	Comedy	country	SA	1526	7.94	1.95	other	3789	7.41	1.84	9.37	0.00	**
105	Adventure	country	SA	580	8.16	1.76	other	1715	7.39	1.78	8.99	0.00	**
120	Adventure	language	ar	506	8.14	1.79	other	1789	7.43	1.78	7.88	0.00	**
18	Action	language	ar	1636	7.59	2.01	other	7006	7.18	1.83	7.67	0.00	**
3	Action	country	SA	1876	7.53	1.97	other	6766	7.18	1.84	6.91	0.00	**
154	Biography	language	ar	649	8.24	1.68	other	1882	7.83	1.45	5.55	0.00	**
87	Comedy	dimension	MA	2290	7.72	1.91	other	3025	7.44	1.86	5.49	0.00	**
89	Comedy	language	ar	1304	7.81	1.97	other	4011	7.48	1.85	5.26	0.00	**
139	Biography	country	SA	779	8.17	1.61	other	1752	7.83	1.48	4.94	0.00	**
14	Action	country	TH	63	8.01	1.32	other	8579	7.25	1.87	4.53	0.00	**
118	Adventure	dimension	MA	983	7.73	1.86	other	1312	7.47	1.76	3.37	0.00	**
88	Comedy	dimension	RE	200	7.95	1.88	other	5115	7.54	1.88	2.95	0.00	**
85	Comedy	country	TH	20	8.75	1.16	other	5295	7.55	1.89	2.81	0.00	**
51	Drama	dimension	MA	5133	7.83	1.88	other	5160	7.73	1.76	2.79	0.01	**
50	Drama	dimension	LA	4852	7.73	1.77	other	5441	7.83	1.87	-2.84	0.00	**
26	Action	language	it	79	6.74	1.48	other	8563	7.26	1.87	-3.08	0.00	**

Table I.1: continued on the next page

Test ID	Genre	Locality	Group1	N1	μ_1	σ_1	Group2	N2	μ_2	σ_1	t-stat	p-value	significance
19	Action	language	en	6190	7.21	1.83	other	2452	7.36	1.96	-3.08	0.00	**
45	Drama	country	GR	43	6.67	2.33	other	10250	7.79	1.82	-3.13	0.00	**
137	Biography	country	GB	362	7.74	1.22	other	2169	7.97	1.57	-3.19	0.00	**
54	Drama	language	en	6483	7.74	1.76	other	3810	7.87	1.92	-3.40	0.00	**
41	Drama	country	IT	153	7.28	1.57	other	10140	7.79	1.83	-3.46	0.00	**
1	Action	country	GB	1525	7.11	1.71	other	7117	7.29	1.90	-3.55	0.00	**
61	Drama	language	it	103	7.13	1.42	other	10190	7.79	1.83	-3.64	0.00	**
97	Comedy	language	it	71	6.90	1.42	other	5244	7.57	1.89	-3.92	0.00	**
20	Action	language	tr	344	6.92	1.58	other	8298	7.27	1.88	-3.97	0.00	**
71	Comedy	country	US	1942	7.42	1.91	other	3373	7.64	1.87	-4.11	0.00	**
36	Drama	country	GB	1433	7.62	1.54	other	8860	7.81	1.86	-4.12	0.00	**
95	Comedy	language	sv	23	5.91	2.33	other	5292	7.57	1.88	-4.20	0.00	**
91	Comedy	language	tr	244	7.11	1.69	other	5071	7.58	1.89	-4.21	0.00	**
22	Action	language	ru	237	6.68	2.10	other	8405	7.27	1.86	-4.25	0.00	**
4	Action	country	TR	534	6.96	1.62	other	8108	7.27	1.88	-4.27	0.00	**
73	Comedy	country	SE	65	6.50	2.12	other	5250	7.57	1.88	-4.54	0.00	**
75	Comedy	country	TR	374	7.14	1.66	other	4941	7.59	1.90	-4.93	0.00	**
6	Action	country	IT	172	6.60	1.58	other	8470	7.27	1.87	-5.47	0.00	**
106	Adventure	country	TR	194	6.96	1.61	other	2101	7.64	1.81	-5.53	0.00	**

Table I.1: continued on the next page

Test ID	Genre	Locality	Group1	N1	μ_1	σ_1	Group2	N2	μ_2	σ_1	t-stat	p-value	significance
140	Biography	country	TR	225	7.37	1.46	other	2306	7.99	1.52	-5.85	0.00	**
156	Biography	language	tr	139	7.20	1.48	other	2392	7.98	1.52	-5.89	0.00	**
5	Action	country	RU	360	6.59	2.00	other	8282	7.28	1.86	-6.46	0.00	**
86	Comedy	dimension	LA	2825	7.40	1.85	other	2490	7.74	1.91	-6.58	0.00	**
122	Adventure	language	tr	122	6.66	1.48	other	2173	7.64	1.81	-6.98	0.00	**
39	Drama	country	TR	922	7.22	1.64	other	9371	7.84	1.83	-9.88	0.00	**
55	Drama	language	tr	623	7.05	1.64	other	9670	7.83	1.82	-10.34	0.00	**
12	Action	country	IN	163	7.60	1.78	other	8479	7.25	1.87	2.35	0.02	*
49	Drama	country	TH	37	8.48	1.23	other	10256	7.78	1.82	2.33	0.02	*
83	Comedy	country	IN	74	8.06	2.02	other	5241	7.55	1.88	2.31	0.02	*
0	Action	country	US	3279	7.31	1.92	other	5363	7.22	1.84	2.16	0.03	*
17	Action	dimension	RE	486	7.43	1.72	other	8156	7.24	1.88	2.12	0.03	*
155	Biography	language	en	1647	7.89	1.43	other	884	8.02	1.68	-2.00	0.05	*
119	Adventure	dimension	RE	94	7.22	1.70	other	2201	7.60	1.81	-2.00	0.05	*
40	Drama	country	RU	199	7.51	1.66	other	10094	7.79	1.83	-2.11	0.03	*
104	Adventure	country	SE	46	6.97	1.90	other	2249	7.60	1.80	-2.32	0.02	*
59	Drama	language	sv	32	6.71	2.59	other	10261	7.79	1.82	-2.34	0.03	*
90	Comedy	language	en	3565	7.52	1.85	other	1750	7.65	1.95	-2.35	0.02	*
2	Action	country	SE	106	6.89	1.57	other	8536	7.26	1.87	-2.38	0.02	*

Table I.1: continued on the next page

Test ID	Genre	Locality	Group1	N1	μ_1	σ_1	Group2	N2	μ_2	σ_1	t-stat	p-value	significance
77	Comedy	country	IT	111	7.23	1.45	other	5204	7.57	1.89	-2.39	0.02	*
127	Adventure	language	es	13	6.38	1.55	other	2282	7.59	1.80	-2.41	0.02	*
37	Drama	country	SE	93	7.33	2.02	other	10200	7.79	1.82	-2.41	0.02	*
72	Comedy	country	GB	818	7.43	1.65	other	4497	7.58	1.92	-2.44	0.01	*
117	Adventure	dimension	LA	1218	7.49	1.76	other	1077	7.69	1.85	-2.55	0.01	*
121	Adventure	language	en	1532	7.51	1.75	other	763	7.73	1.90	-2.56	0.01	*
153	Biography	dimension	RE	73	8.27	1.19	other	2458	7.92	1.53	1.90	0.06	
157	Biography	language	ru	27	8.44	1.78	other	2504	7.93	1.52	1.73	0.08	
58	Drama	language	th	2	10.00	0.00	other	10291	7.78	1.82	1.71	0.09	
93	Comedy	language	ru	69	7.92	1.68	other	5246	7.55	1.89	1.61	0.11	
148	Biography	country	IN	19	8.42	1.01	other	2512	7.93	1.53	1.38	0.17	
150	Biography	country	TH	10	8.60	0.84	other	2521	7.93	1.53	1.37	0.17	
23	Action	language	th	9	8.11	1.61	other	8633	7.25	1.87	1.36	0.17	
136	Biography	country	US	867	7.99	1.56	other	1664	7.91	1.51	1.31	0.19	
82	Comedy	country	ID	80	7.80	1.71	other	5235	7.56	1.89	1.13	0.26	
147	Biography	country	ID	35	8.20	1.43	other	2496	7.93	1.53	1.02	0.31	
101	Comedy	language	ca	3	8.66	1.52	other	5312	7.56	1.89	1.01	0.31	
56	Drama	language	zh	4	8.50	1.00	other	10289	7.78	1.82	0.78	0.44	
7	Action	country	BR	111	7.39	1.89	other	8531	7.25	1.87	0.77	0.44	

Table I.1: continued on the next page

Test ID	Genre	Locality	Group1	N1	μ_1	σ_1	Group2	N2	μ_2	σ_1	t-stat	p-value	significance
67	Drama	language	da	6	8.33	1.50	other	10287	7.78	1.82	0.73	0.47	
80	Comedy	country	VE	37	7.78	1.54	other	5278	7.56	1.89	0.71	0.48	
92	Comedy	language	zh	2	8.50	2.12	other	5313	7.56	1.89	0.70	0.48	
47	Drama	country	IN	87	7.91	1.85	other	10206	7.78	1.82	0.67	0.50	
24	Action	language	sv	42	7.38	1.43	other	8600	7.25	1.87	0.55	0.58	
28	Action	language	fr	9	7.55	1.01	other	8633	7.25	1.87	0.47	0.64	
13	Action	country	JP	63	7.36	1.57	other	8579	7.25	1.87	0.45	0.65	
132	Adventure	language	pt	7	7.85	1.57	other	2288	7.58	1.81	0.39	0.70	
33	Action	language	id	3	7.66	0.57	other	8639	7.25	1.87	0.38	0.71	
146	Biography	country	GR	8	8.12	1.35	other	2523	7.93	1.53	0.34	0.73	
79	Comedy	country	ES	50	7.64	2.11	other	5265	7.56	1.88	0.29	0.78	
152	Biography	dimension	MA	1200	7.94	1.58	other	1331	7.93	1.47	0.23	0.82	
29	Action	language	pt	24	7.33	1.83	other	8618	7.25	1.87	0.19	0.85	
141	Biography	country	RU	49	7.97	1.63	other	2482	7.93	1.52	0.19	0.85	
43	Drama	country	ES	141	7.81	1.78	other	10152	7.78	1.82	0.18	0.86	
35	Drama	country	US	3326	7.79	1.85	other	6967	7.78	1.81	0.13	0.90	
52	Drama	dimension	RE	308	7.79	1.72	other	9985	7.78	1.83	0.10	0.92	
48	Drama	country	JP	51	7.80	1.85	other	10242	7.78	1.82	0.06	0.95	
70	Drama	language	id	5	7.80	0.44	other	10288	7.78	1.82	0.06	0.96	

Table I.1: continued on the next page

Test ID	Genre	Locality	Group1	N1	μ_1	σ_1	Group2	N2	μ_2	σ_1	t-stat	p-value	significance
25	Action	language	es	25	7.24	1.92	other	8617	7.25	1.87	-0.05	0.96	
149	Biography	country	JP	9	7.88	0.78	other	2522	7.93	1.53	-0.10	0.92	
16	Action	dimension	MA	3246	7.25	1.92	other	5396	7.26	1.84	-0.13	0.90	
100	Comedy	language	pt	16	7.50	2.78	other	5299	7.56	1.88	-0.14	0.89	
69	Drama	language	he	2	7.50	0.70	other	10291	7.78	1.82	-0.22	0.82	
8	Action	country	ES	122	7.22	1.78	other	8520	7.25	1.87	-0.23	0.82	
116	Adventure	country	TH	9	7.44	1.42	other	2286	7.59	1.81	-0.24	0.81	
63	Drama	language	fr	8	7.62	2.44	other	10285	7.78	1.82	-0.25	0.80	
84	Comedy	country	JP	26	7.46	2.26	other	5289	7.56	1.88	-0.28	0.78	
76	Comedy	country	RU	113	7.51	1.99	other	5202	7.56	1.88	-0.29	0.77	
114	Adventure	country	IN	25	7.48	1.85	other	2270	7.59	1.80	-0.30	0.76	
96	Comedy	language	es	11	7.36	1.80	other	5304	7.56	1.89	-0.35	0.72	
164	Biography	language	pt	12	7.75	1.35	other	2519	7.94	1.53	-0.43	0.67	
62	Drama	language	ja	13	7.53	2.06	other	10280	7.78	1.82	-0.49	0.62	
107	Adventure	country	RU	69	7.44	2.41	other	2226	7.59	1.78	-0.49	0.62	
124	Adventure	language	ru	53	7.39	2.67	other	2242	7.59	1.78	-0.54	0.59	
66	Drama	language	ca	11	7.63	0.80	other	10282	7.78	1.82	-0.62	0.55	
159	Biography	language	es	6	7.50	0.54	other	2525	7.94	1.53	-0.70	0.48	
44	Drama	country	VE	66	7.62	1.53	other	10227	7.78	1.83	-0.75	0.46	

Table I.1: continued on the next page

Test ID	Genre	Locality	Group1	N1	μ_1	σ_1	Group2	N2	μ_2	σ_1	t-stat	p-value	significance
110	Adventure	country	ES	25	7.32	1.77	other	2270	7.59	1.81	-0.75	0.45	
109	Adventure	country	BR	34	7.35	1.85	other	2261	7.59	1.80	-0.77	0.44	
78	Comedy	country	BR	69	7.37	2.12	other	5246	7.56	1.88	-0.83	0.41	
112	Adventure	country	GR	20	7.25	1.86	other	2275	7.59	1.80	-0.84	0.40	
42	Drama	country	BR	137	7.65	1.68	other	10156	7.79	1.83	-0.85	0.40	
143	Biography	country	BR	49	7.75	1.40	other	2482	7.94	1.53	-0.85	0.40	
133	Adventure	language	ca	2	6.50	0.70	other	2293	7.59	1.81	-0.85	0.39	
15	Action	dimension	LA	4910	7.24	1.85	other	3732	7.27	1.90	-0.86	0.39	
165	Biography	language	ca	2	7.00	1.41	other	2529	7.93	1.53	-0.87	0.39	
151	Biography	dimension	LA	1258	7.91	1.49	other	1273	7.96	1.56	-0.87	0.38	
102	Adventure	country	US	772	7.54	1.87	other	1523	7.61	1.77	-0.93	0.35	
126	Adventure	language	sv	22	7.22	1.87	other	2273	7.59	1.80	-0.94	0.35	
11	Action	country	ID	197	7.13	1.77	other	8445	7.26	1.87	-0.96	0.34	
10	Action	country	GR	25	6.88	2.08	other	8617	7.26	1.87	-1.01	0.31	
27	Action	language	ja	26	6.84	1.73	other	8616	7.26	1.87	-1.12	0.26	
138	Biography	country	SE	29	7.62	1.95	other	2502	7.94	1.52	-1.13	0.26	
98	Comedy	language	ja	5	6.60	3.28	other	5310	7.56	1.88	-1.14	0.25	
144	Biography	country	ES	36	7.63	1.35	other	2495	7.94	1.53	-1.19	0.24	
65	Drama	language	pt	35	7.40	1.81	other	10258	7.78	1.82	-1.26	0.21	

Table I.1: continued on the next page

Test ID	Genre	Locality	Group1	N1	μ_1	σ_1	Group2	N2	μ_2	σ_1	t-stat	p-value	significance
81	Comedy	country	GR	10	6.80	1.47	other	5305	7.56	1.89	-1.28	0.20	
57	Drama	language	ru	126	7.57	1.58	other	10167	7.79	1.83	-1.29	0.20	
60	Drama	language	es	25	7.28	1.40	other	10268	7.78	1.82	-1.39	0.16	
64	Drama	language	el	4	6.50	1.91	other	10289	7.78	1.82	-1.41	0.16	
9	Action	country	VE	46	6.86	1.68	other	8596	7.26	1.87	-1.41	0.16	
21	Action	language	zh	3	5.66	0.57	other	8639	7.25	1.87	-1.47	0.14	
103	Adventure	country	GB	400	7.48	1.51	other	1895	7.61	1.86	-1.52	0.13	
115	Adventure	country	JP	19	6.94	1.77	other	2276	7.59	1.80	-1.55	0.12	
113	Adventure	country	ID	41	7.14	1.66	other	2254	7.59	1.81	-1.58	0.11	
30	Action	language	ca	11	6.36	1.62	other	8631	7.26	1.87	-1.58	0.11	
46	Drama	country	ID	133	7.52	1.65	other	10160	7.79	1.83	-1.66	0.10	
31	Action	language	da	2	5.00	1.41	other	8640	7.25	1.87	-1.70	0.09	
129	Adventure	language	ja	10	6.60	1.71	other	2285	7.59	1.80	-1.73	0.08	
128	Adventure	language	it	22	6.90	1.47	other	2273	7.59	1.81	-1.77	0.08	
142	Biography	country	IT	38	7.50	1.17	other	2493	7.94	1.53	-1.78	0.07	
160	Biography	language	it	31	7.45	1.20	other	2500	7.94	1.53	-1.79	0.07	
145	Biography	country	VE	16	7.25	1.18	other	2515	7.94	1.53	-1.81	0.07	
108	Adventure	country	IT	38	7.05	1.67	other	2257	7.59	1.81	-1.84	0.07	
111	Adventure	country	VE	23	6.86	1.79	other	2272	7.59	1.80	-1.92	0.06	

Table I.1: continued on the next page

Test ID	Genre	Locality	Group1	N1	μ_1	σ_1	Group2	N2	μ_2	σ_1	t-stat	p-value	significance
158	Biography	language	sv	14	7.14	2.17	other	2517	7.94	1.52	-1.95	0.05	
32	Action	language	he	1	5.00	-	other	8641	7.25	1.87	-	-	
34	Action	language	nl	1	9.00	-	other	8641	7.25	1.87	-	-	
68	Drama	language	pl	1	9.00	-	other	10292	7.78	1.82	-	-	
94	Comedy	language	th	1	9.00	-	other	5314	7.56	1.89	-	-	
99	Comedy	language	fr	1	9.00	-	other	5314	7.56	1.89	-	-	
123	Adventure	language	zh	1	7.00	-	other	2294	7.58	1.81	-	-	
125	Adventure	language	th	1	7.00	-	other	2294	7.58	1.81	-	-	
130	Adventure	language	fr	1	7.00	-	other	2294	7.58	1.81	-	-	
131	Adventure	language	el	1	8.00	-	other	2294	7.58	1.81	-	-	
134	Adventure	language	da	1	9.00	-	other	2294	7.58	1.81	-	-	
135	Adventure	language	nl	1	7.00	-	other	2294	7.58	1.81	-	-	
161	Biography	language	ja	1	9.00	-	other	2530	7.93	1.53	-	-	
162	Biography	language	fr	1	7.00	-	other	2530	7.93	1.53	-	-	
163	Biography	language	el	1	8.00	-	other	2530	7.93	1.53	-	-	
166	Biography	language	da	1	8.00	-	other	2530	7.93	1.53	-	-	

Table I.1: Comparison of genre ratings of user country groups

Appendix J

Recommendation System

All manual and offline experiments were performed on MAC OS X El Capitan computer with CPU 3,4 GHz Intel Core i7 and 32GB memory.

J.1 Manual testing

In this section we provide the manual testing results of the developed recommender system. The manual testing was performed using "SAMPLE" dataset described below and constituting 7000 movie ratings . Besides the results described in the thesis body, it was important to include the pilot tests working with LASSO ("Linear Model trained with L1 prior as regularizer") and ("Random Forest") models realised by scikit-learn.org. Both models were excluded from the further tests due to poor performance of the LASSO model in NDCG and RMSE metrics. The random Forest was excluded mainly due greater computation time, and also considerable rating prediction error. Additionally, learning curves and feature rankings and performance improvement cut-off summaries in individual offline user tests are provided.

Data Table	SAMPLE	TIMELINE	ALL
ratings	7000	32596	39596
users	3001	6349	6891
movies	2411	6711	7446
countries	15	15	15
languages	17	20	20
mean rating	7.64	7.62	7.63
std. rating	1.84	1.85	1.85

Table J.1: Rating data tables for recommender system tests: extended Information (data collection from 2017-01-01 to 2017-05-27

J.1.1 Summary of Rating Tables

Table J.1 above shows the extended information on rating data tables used in automated offline and manual testing.

J.1.2 Country Inference Sample

Table J.2 rows show particular data samples with user's country and language provided by Twitter API, free-text location field (typed in by a user in one's profile) and user time zone. The last column shows the inferred country when considering PLACE feature set (combining user language, location and time zone).

provided country	language	location	time zone	inferred country
GB	en	Clacton-on-sea, Essex		GB
US	en	Northern Cali		US
GB	en	Stafford London		GB
US	ar	Where I meant to be.	Baghdad	US
US	en	Sherman Oaks	Pacific Time (US & Canada)	US
GB	en	London		GB
US	en	412 Central Time	(US & Canada)	US
RU	en	russia, barnaul	Novosibirsk	RU

Table J.2: Country inference sample using PLACE feature set (language, free-text location and time zone)

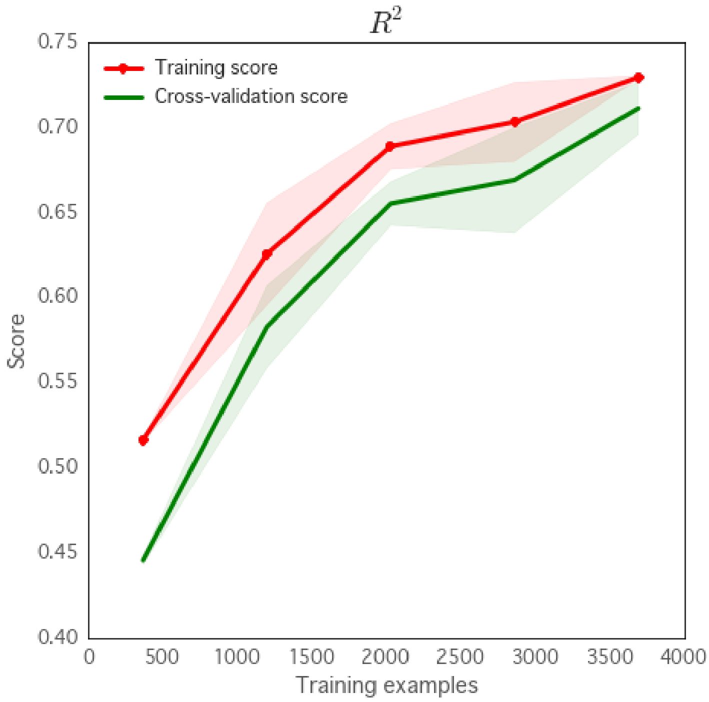


Figure J.1: Learning curve for FACTORS model

J.1.3 Learning Curves

In this section we demonstrate learning curves for three main recommendation approaches we further tested (FACTORS for factorisation machines, OFFSET based on user and movie average ratings, and BOOSTER for gradient boosting regression) tests on all features available ("ALL" feature set). Additionally, a Random Forest Regression ("FOREST"), which, however, we disregard in further tests due to substantial training time demands when tuning and training using larger data sample. The learning curves show the R^2 performance when the related machine learning approach is tested on training and cross-validation datasets. Figures J.1, J.2 and J.3 for FACTORS, OFFSET and BOOSTER models respectively, indicating that the OFFSET model is limited to R^2 cross-validation score less than 70%, decreasing while adding more training instances, while BOOSTER model R^2 cross-validation score can reach 75% with about 1200 training instances. The FACTORS model shows relatively poor cross-validation R^2 score while having few training instances, which, however, demonstrates a very rapid performance improvement trend as compared with OFFSET and BOOSTER approaches.

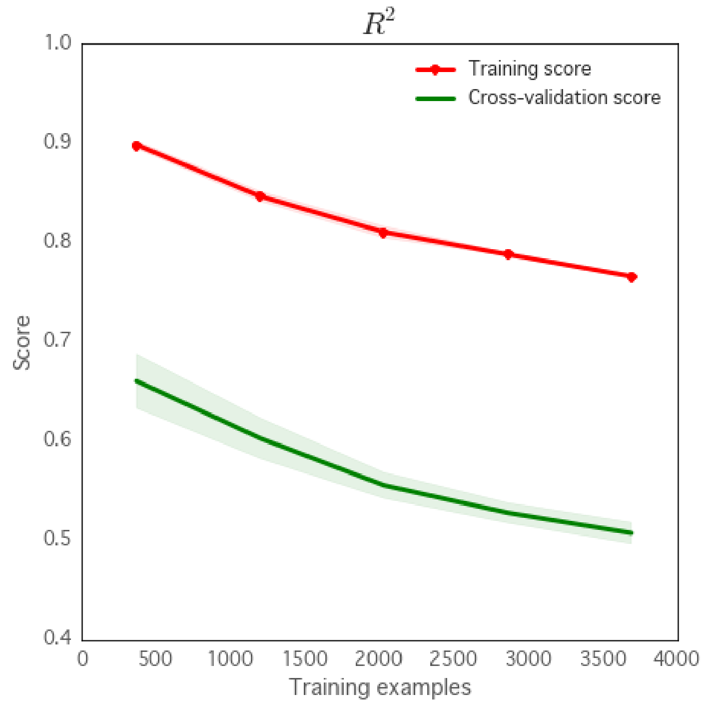


Figure J.2: Learning curve for OFFSET model

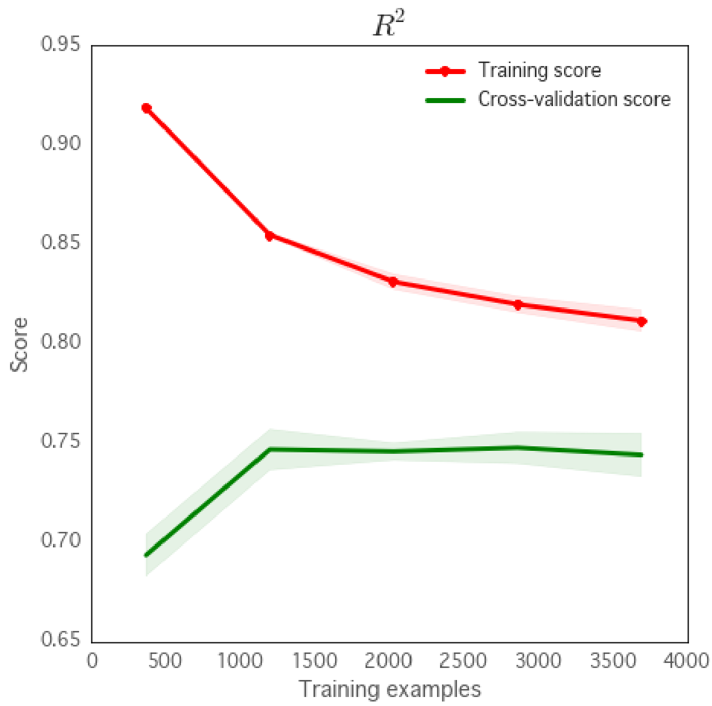


Figure J.3: Learning curve for BOOSTER model

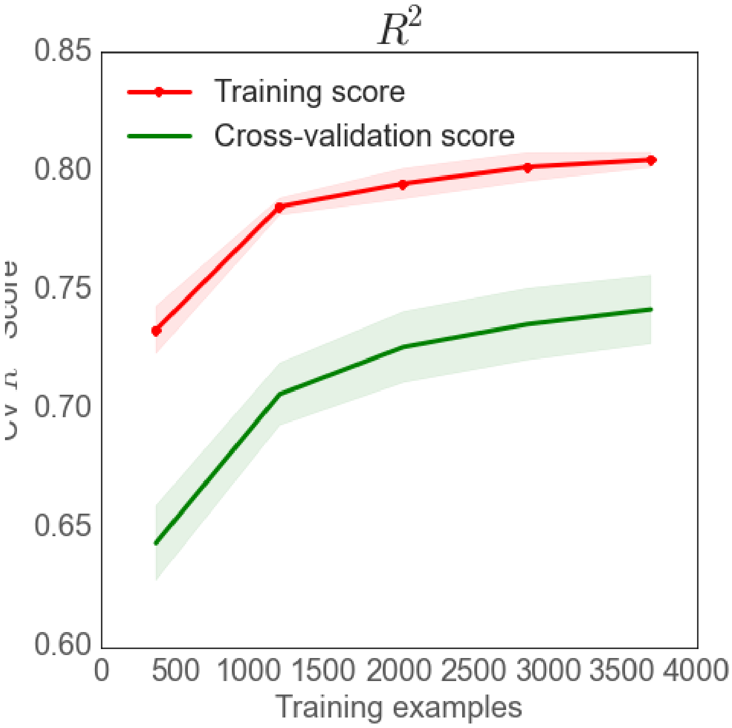


Figure J.4: Learning curve for FOREST (random forest regression) model

J.1.4 Feature Importance

Figure J.5 shows the relative importance of the top 15 features calculated using BOOSTER (gradient boosting regression) and FOREST (Random Forest Regression). The user average movie rating is the top feature in the top 15 most important features when both models are trained using small "SAMPLE" training dataset. We observed that both models give the highest importance to user and movie average ratings. More tests are required whether exploiting additional characteristics would lead to improved performance. Please refer to the next Appendix J.2 for offline tests considering different context-inclusion strategies, pre-filtering, and "timeline" tests.

J.1.5 Performance for Context Variables Usage in Pilot Tests

Before performing tests on the larger dataset, we evaluated several machine learning approaches to create transaction-based recommendation systems. For each of the machine learning techniques, we exploited different feature sets called "context inclusion strategies". Figures J.6 and J.7 show pilot test performance of different recommendation models and context inclusion strategies for RMSE (rating predic-

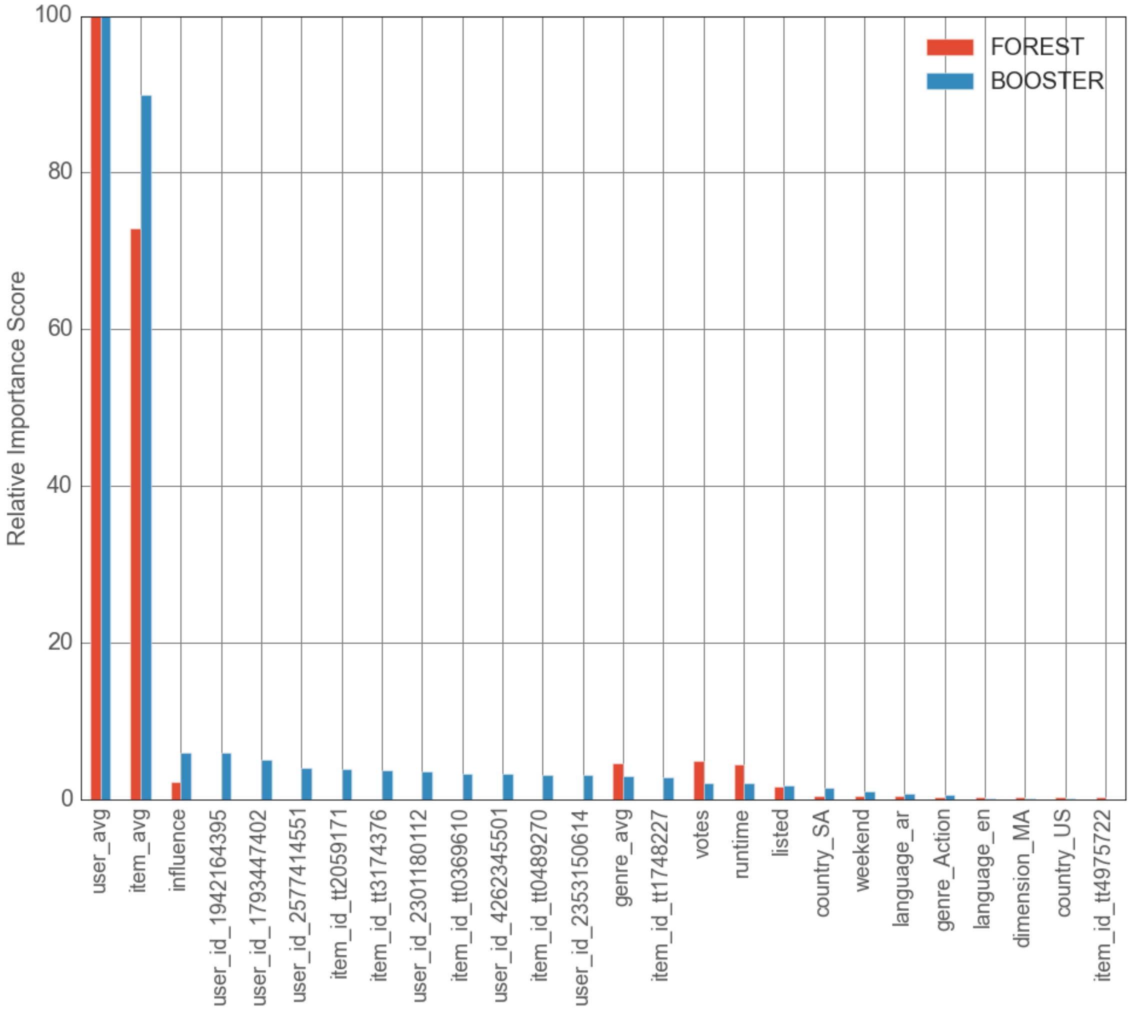


Figure J.5: Features importance for BOOSTER (gradient boosting regression) and FOREST (random forest) models

tion error) and NDCG (ranking performance) metrics respectively. Figures J.6 indicates that the OFFSET, FACTORS, and BOOSTER could enable the smallest and comparable RMSE performance when tested on the “SAMPLE” dataset across all feature sets. The ranking performance in NDCG metric was the best for FACTORS when using inferred COUNTRY feature set (all averages and inferred country), LOCALITY feature set (all averages and inferred country, cultural dimension and user language defined in the Twitter profile), and also 2AVERAGES (when only user and movie average ratings were used). The BOOSTER model, however, benefited from adding the average genre rating for the inferred country when using 3AVERAGES feature set.

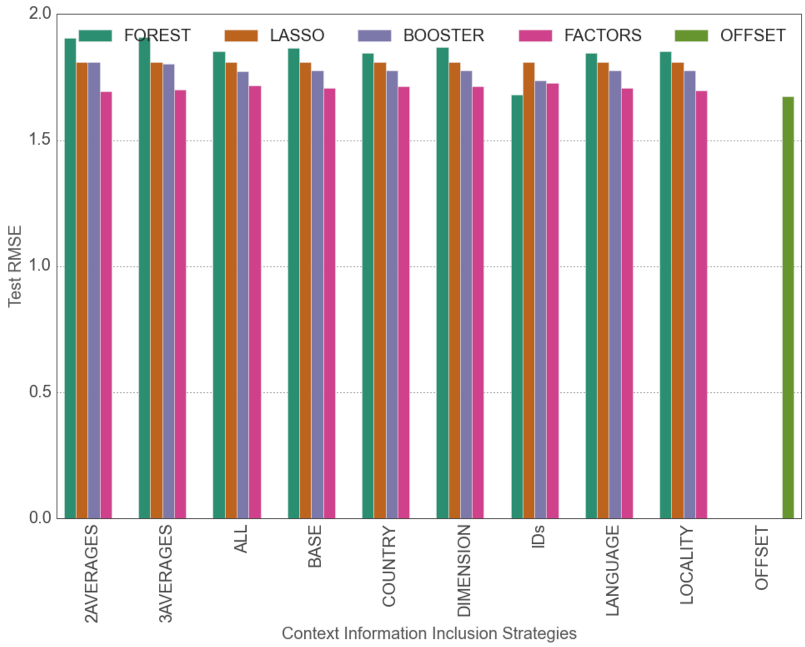


Figure J.6: RMSE performance in pilot tests

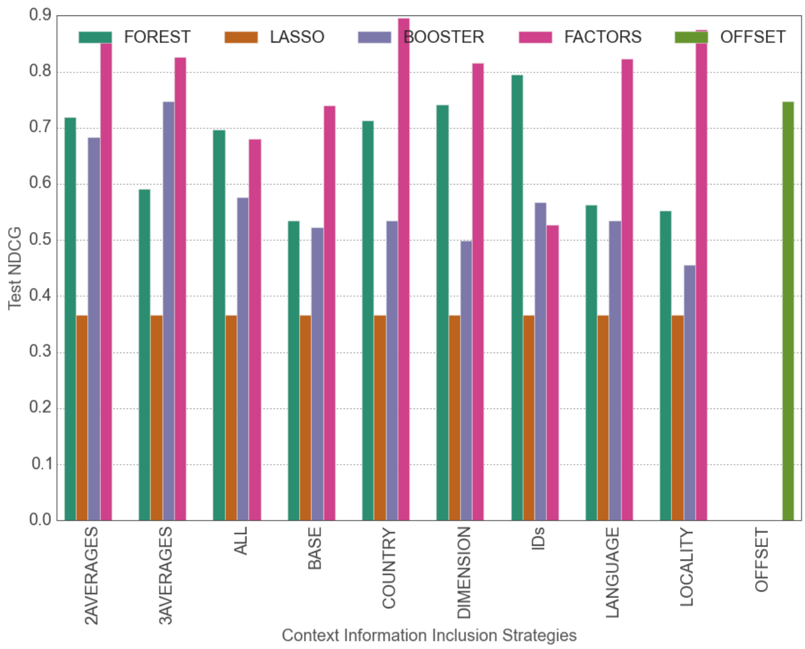


Figure J.7: NDCG ranking performance in pilot tests (we also included two additional feature sets including “2AVERAGES”, composed of user and movie average rating, and “3AVERAGES”, in which we also considered an average rating of the associated with movie top genre and the inferred country)

J.2 Offline Tests

J.2.1 Ranking context-aware Recommendation Strategies

	R^2	NDCG	RMSE	Overall Rank
BOOSTER:ALL	18.00	16.00	17.50	17.17
BOOSTER:BASE	15.00	14.00	16.50	15.17
BOOSTER:LOCALITY	14.00	19.00	15.50	16.17
BOOSTER:COUNTRY	19.00	18.00	14.50	17.17
BOOSTER:DIMENSION	16.00	17.00	13.50	15.50
BOOSTER:LANGUAGE	17.00	15.00	12.50	14.83
FACTORS:IDs	4.00	3.00	11.50	6.17
FACTORS:DIMENSION	11.00	7.00	10.50	9.50
FACTORS:BASE	9.00	9.00	9.50	9.17
FACTORS:COUNTRY	10.00	12.00	8.50	10.17
FACTORS:LOCALITY	8.00	8.00	7.50	7.83
OFFSET:IDs	7.00	13.00	6.50	8.83
FACTORS:LANGUAGE	12.00	11.00	5.50	9.50
ITEM:IDs	6.00	4.00	4.50	4.83
FACTORS:ALL	13.00	6.00	3.50	7.50
BOOSTER:IDs	3.00	5.00	2.50	3.50
USER:IDs	5.00	10.00	1.50	5.50
MU:IDs	1.50	1.50	0.00	1.00

Table J.3: Ranking context-aware recommendation strategies

J.2.2 Pre-filtering and Context Inclusion

"Baseline" and "Tested" columns indicate recommendation strategies in a form "Case:Model:Context". "Case" includes "REG" for regular recommendation state when user ratings available in train and test sets, "FILT" when we pre-filter the whole dataset in accord with the user cultural dimension, "COLD" when no ratings data is available in the user train dataset, "COLD_FILT" when we pre-filter the whole dataset in accord with user cultural dimension while having cold-start situation. Model includes "OFFSET" based on user and movie average ratings and offsets, "FACTORS" using Factorisation Machines by [199], and "BOOSTER" based on the Gradient Boosting Regression model. The "Context" feature sets are described in Table 9.3.

#	case	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
51	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:BOOSTER:IDs	1.33	0.88	33.07	4.70	< 0.01	164	-
44	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:FACTORS:IDs	1.30	0.83	29.65	4.48	< 0.01	164	-
8	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:BOOSTER:IDs	1.28	0.90	22.64	5.80	< 0.01	204	-
1	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:FACTORS:IDs	1.19	0.89	13.75	4.19	< 0.01	204	-
43	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:FACTORS:ALL	1.14	0.73	13.49	2.08	< 0.05	164	-
45	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:FACTORS:BASE	1.13	0.70	12.91	2.12	< 0.05	164	-
48	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:FACTORS:LANGUAGE	1.12	0.73	11.37	1.74	0.08	164	~
42	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:OFFSET:ALL	1.09	0.76	8.18	1.29	0.20	164	~
50	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:BOOSTER:ALL	1.07	0.72	6.66	1.04	0.30	164	~
49	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:FACTORS:LOCALITY	1.07	0.72	6.40	1.01	0.31	164	~
46	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:FACTORS:COUNTRY	1.06	0.70	6.03	1.00	0.32	164	~
47	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:FACTORS:DIMENSION	1.05	0.70	4.31	0.72	0.47	164	~
55	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:BOOSTER:LANGUAGE	1.02	0.71	1.96	0.32	0.75	164	~
56	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:BOOSTER:LOCALITY	1.02	0.71	1.68	0.28	0.78	164	~
52	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:BOOSTER:BASE	1.02	0.71	1.32	0.22	0.83	164	~
54	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:BOOSTER:DIMENSION	1.02	0.70	1.22	0.20	0.84	164	~
53	FILT	RMSE	REG:OFFSET:ALL	1.00	0.69	FILT:BOOSTER:COUNTRY	1.01	0.70	0.42	0.07	0.94	164	~

Table J.4: continued on the next page

#	case	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
7	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:BOOSTER:ALL	1.00	0.82	-4.17	-1.29	0.20	204	~
6	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:FACTORS:LOCALITY	0.99	0.79	-5.34	-1.75	0.08	204	~
0	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:FACTORS:ALL	0.98	0.73	-5.66	-1.85	0.07	204	~
12	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:BOOSTER:LANGUAGE	0.98	0.80	-6.47	-2.03	< 0.05	204	+
11	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:BOOSTER:DIMENSION	0.98	0.80	-6.53	-2.03	< 0.05	204	+
4	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:FACTORS:DIMENSION	0.97	0.73	-6.58	-2.21	< 0.05	204	+
10	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:BOOSTER:COUNTRY	0.97	0.79	-7.00	-2.17	< 0.05	204	+
5	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:FACTORS:LANGUAGE	0.97	0.74	-7.18	-2.38	< 0.05	204	+
13	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:BOOSTER:LOCALITY	0.97	0.77	-7.32	-2.31	< 0.05	204	+
3	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:FACTORS:COUNTRY	0.97	0.73	-7.39	-2.48	< 0.05	204	+
9	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:BOOSTER:BASE	0.96	0.77	-7.62	-2.47	< 0.05	204	+
2	REG	RMSE	REG:OFFSET:ALL	1.04	0.81	REG:FACTORS:BASE	0.96	0.74	-8.45	-2.87	< 0.01	204	+
57	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:OFFSET:ALL	0.96	0.07	1.88	2.15	< 0.05	164	+
68	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:BOOSTER:COUNTRY	0.96	0.08	1.82	2.42	< 0.05	164	+
61	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:FACTORS:COUNTRY	0.96	0.08	1.81	2.54	< 0.05	164	+
62	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:FACTORS:DIMENSION	0.96	0.08	1.74	2.39	< 0.05	164	+
63	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:FACTORS:LANGUAGE	0.96	0.08	1.69	2.29	< 0.05	164	+
60	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:FACTORS:BASE	0.96	0.08	1.60	2.14	< 0.05	164	+

Table J.4: continued on the next page

#	case	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
70	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:BOOSTER:LANGUAGE	0.96	0.08	1.54	1.84	0.07	164	~
71	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:BOOSTER:LOCALITY	0.96	0.08	1.53	2.01	< 0.05	164	+
67	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:BOOSTER:BASE	0.96	0.08	1.51	1.98	< 0.05	164	+
69	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:BOOSTER:DIMENSION	0.96	0.08	1.48	1.89	0.06	164	~
64	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:FACTORS:LOCALITY	0.96	0.08	1.44	1.79	0.08	164	~
58	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:FACTORS:ALL	0.96	0.08	1.42	1.62	0.11	164	~
65	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:BOOSTER:ALL	0.96	0.08	1.17	1.40	0.16	164	~
17	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:FACTORS:COUNTRY	0.96	0.09	0.85	1.64	0.10	204	~
19	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:FACTORS:LANGUAGE	0.96	0.09	0.69	1.37	0.17	204	~
21	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:BOOSTER:ALL	0.96	0.09	0.56	1.09	0.28	204	~
16	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:FACTORS:BASE	0.96	0.09	0.47	0.94	0.35	204	~
23	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:BOOSTER:BASE	0.96	0.09	0.45	0.93	0.35	204	~
27	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:BOOSTER:LOCALITY	0.96	0.09	0.45	0.93	0.35	204	~
20	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:FACTORS:LOCALITY	0.96	0.09	0.43	0.85	0.40	204	~
14	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:FACTORS:ALL	0.96	0.09	0.40	0.80	0.43	204	~
26	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:BOOSTER:LANGUAGE	0.96	0.09	0.40	0.81	0.42	204	~
25	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:BOOSTER:DIMENSION	0.95	0.09	0.37	0.76	0.45	204	~
18	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:FACTORS:DIMENSION	0.95	0.09	0.37	0.76	0.45	204	~

Table J.4: continued on the next page

#	case	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
24	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:BOOSTER:COUNTRY	0.95	0.09	0.31	0.63	0.53	204	~
59	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:FACTORS:IDs	0.94	0.11	-1.10	-1.09	0.28	164	~
66	FILT	NDCG	REG:OFFSET:ALL	0.95	0.10	FILT:BOOSTER:IDs	0.93	0.11	-1.62	-1.53	0.13	164	~
15	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:FACTORS:IDs	0.93	0.11	-1.80	-2.88	< 0.01	204	-
22	REG	NDCG	REG:OFFSET:ALL	0.95	0.09	REG:BOOSTER:IDs	0.93	0.11	-1.88	-2.67	< 0.01	204	-
72	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:OFFSET:ALL	0.48	0.05	11.04	13.33	< 0.01	164	+
41	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:BOOSTER:LOCALITY	0.48	0.00	10.19	inf	< 0.01	204	+
37	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:BOOSTER:BASE	0.48	0.00	9.83	inf	< 0.01	204	+
30	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:FACTORS:BASE	0.47	0.00	8.52	inf	< 0.01	204	+
38	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:BOOSTER:COUNTRY	0.47	0.00	8.17	inf	< 0.01	204	+
39	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:BOOSTER:DIMENSION	0.46	0.00	7.17	inf	< 0.01	204	+
40	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:BOOSTER:LANGUAGE	0.46	0.00	6.69	inf	< 0.01	204	+
31	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:FACTORS:COUNTRY	0.46	0.00	5.22	inf	< 0.01	204	+
32	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:FACTORS:DIMENSION	0.45	0.00	3.22	inf	< 0.01	204	+
33	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:FACTORS:LANGUAGE	0.44	0.00	2.73	inf	< 0.01	204	+
35	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:BOOSTER:ALL	0.44	0.00	2.44	inf	< 0.01	204	+
75	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:FACTORS:BASE	0.44	0.05	0.73	0.81	0.42	164	~
77	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:FACTORS:DIMENSION	0.43	0.04	-0.15	-0.19	0.85	164	~

Table J.4: continued on the next page

#	case	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
86	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:BOOSTER:LOCALITY	0.43	0.05	-0.39	-0.40	0.69	164	~
82	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:BOOSTER:BASE	0.43	0.05	-0.92	-1.00	0.32	164	~
83	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:BOOSTER:COUNTRY	0.43	0.05	-1.10	-1.14	0.26	164	~
34	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:FACTORS:LOCALITY	0.42	0.00	-1.87	-inf	< 0.01	204	-
85	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:BOOSTER:LANGUAGE	0.42	0.06	-2.53	-2.52	< 0.05	164	-
84	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:BOOSTER:DIMENSION	0.42	0.06	-2.58	-2.55	< 0.05	164	-
28	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:FACTORS:ALL	0.42	0.00	-4.02	-inf	< 0.01	204	-
76	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:FACTORS:COUNTRY	0.41	0.05	-4.75	-4.84	< 0.01	164	-
80	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:BOOSTER:ALL	0.38	0.06	-11.31	-10.35	< 0.01	164	-
79	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:FACTORS:LOCALITY	0.38	0.03	-12.84	-20.45	< 0.01	164	-
78	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:FACTORS:LANGUAGE	0.38	0.07	-13.07	-10.57	< 0.01	164	-
73	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:FACTORS:ALL	0.36	0.06	-17.87	-16.80	< 0.01	164	-
29	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:FACTORS:IDs	0.25	0.00	-42.91	-inf	< 0.01	204	-
74	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:FACTORS:IDs	0.21	0.07	-52.63	-44.14	< 0.01	164	-
36	REG	R^2	REG:OFFSET:ALL	0.43	0.00	REG:BOOSTER:IDs	0.20	0.00	-52.74	-4663.98	< 0.01	204	-
81	FILT	R^2	REG:OFFSET:ALL	0.43	0.00	FILT:BOOSTER:IDs	0.16	0.08	-63.10	-44.22	< 0.01	164	-
						Cold-Start							
51	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:BOOSTER:IDs	1.54	0.61	41.71	14.94	< 0.01	270	-

Table J.4: continued on the next page

#	case	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
44	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:FACTORS:IDs	1.53	0.61	40.38	14.21	< 0.01	270	-
8	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:BOOSTER:IDs	1.52	0.60	40.20	14.95	< 0.01	270	-
1	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:FACTORS:IDs	1.48	0.61	36.50	13.94	< 0.01	270	-
49	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:FACTORS:LOCALITY	1.27	0.55	17.11	7.96	< 0.01	270	-
6	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:FACTORS:LOCALITY	1.22	0.57	11.97	6.04	< 0.01	270	-
43	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:FACTORS:ALL	1.20	0.51	10.74	6.50	< 0.01	270	-
48	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:FACTORS:LANGUAGE	1.19	0.55	9.61	5.10	< 0.01	270	-
2	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:FACTORS:BASE	1.17	0.52	7.72	4.57	< 0.01	270	-
50	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:BOOSTER:ALL	1.16	0.51	6.67	3.70	< 0.01	270	-
0	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:FACTORS:ALL	1.14	0.52	5.04	3.34	< 0.01	270	-
45	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:FACTORS:BASE	1.13	0.51	4.28	2.78	< 0.01	270	-
7	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:BOOSTER:ALL	1.13	0.54	3.93	2.47	< 0.05	270	-
46	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:FACTORS:COUNTRY	1.13	0.51	3.85	2.68	< 0.01	270	-
54	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:BOOSTER:DIMENSION	1.12	0.50	2.91	1.80	0.07	270	~
55	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:BOOSTER:LANGUAGE	1.12	0.50	2.82	1.72	0.09	270	~
47	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:FACTORS:DIMENSION	1.12	0.51	2.63	1.88	0.06	270	~
53	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:BOOSTER:COUNTRY	1.11	0.49	2.54	1.59	0.11	270	~
52	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:BOOSTER:BASE	1.11	0.51	2.21	1.40	0.16	270	~

Table J.4: continued on the next page

#	case	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
56	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:BOOSTER:LOCALITY	1.11	0.51	2.07	1.32	0.19	270	~
5	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:FACTORS:LANGUAGE	1.11	0.53	1.93	1.46	0.15	270	~
3	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:FACTORS:COUNTRY	1.10	0.52	1.32	1.00	0.32	270	~
4	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:FACTORS:DIMENSION	1.10	0.53	1.25	0.97	0.34	270	~
10	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:BOOSTER:COUNTRY	1.10	0.52	0.91	0.63	0.53	270	~
11	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:BOOSTER:DIMENSION	1.10	0.53	0.86	0.61	0.54	270	~
12	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:BOOSTER:LANGUAGE	1.09	0.52	0.68	0.48	0.63	270	~
13	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:BOOSTER:LOCALITY	1.09	0.52	0.16	0.11	0.91	270	~
9	COLD	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD:BOOSTER:BASE	1.09	0.52	0.07	0.05	0.96	270	~
42	COLD_FILT	RMSE	COLD:OFFSET:ALL	1.09	0.53	COLD_FILT:OFFSET:ALL	1.04	0.52	-3.88	-2.59	< 0.05	270	+
57	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:OFFSET:ALL	0.90	0.11	0.86	1.34	0.18	270	~
63	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:FACTORS:LANGUAGE	0.90	0.11	0.49	0.82	0.41	270	~
62	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:FACTORS:DIMENSION	0.90	0.11	0.48	0.80	0.43	270	~
61	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:FACTORS:COUNTRY	0.90	0.11	0.31	0.51	0.61	270	~
27	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:BOOSTER:LOCALITY	0.90	0.12	0.18	0.32	0.75	270	~
23	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:BOOSTER:BASE	0.90	0.12	0.17	0.30	0.76	270	~
60	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:FACTORS:BASE	0.90	0.11	0.08	0.13	0.90	270	~
18	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:FACTORS:DIMENSION	0.90	0.12	0.02	0.03	0.98	270	~

Table J.4: continued on the next page

#	case	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
19	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:FACTORS:LANGUAGE	0.90	0.12	-0.02	-0.03	0.98	270	~
17	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:FACTORS:COUNTRY	0.89	0.12	-0.23	-0.41	0.68	270	~
71	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:BOOSTER:LOCALITY	0.89	0.12	-0.27	-0.43	0.67	270	~
67	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:BOOSTER:BASE	0.89	0.12	-0.28	-0.46	0.65	270	~
68	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:BOOSTER:COUNTRY	0.89	0.12	-0.29	-0.46	0.64	270	~
25	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:BOOSTER:DIMENSION	0.89	0.13	-0.51	-0.88	0.38	270	~
26	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:BOOSTER:LANGUAGE	0.89	0.13	-0.57	-0.98	0.33	270	~
70	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:BOOSTER:LANGUAGE	0.89	0.12	-0.63	-0.97	0.33	270	~
69	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:BOOSTER:DIMENSION	0.89	0.12	-0.63	-0.97	0.33	270	~
14	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:FACTORS:ALL	0.89	0.13	-0.74	-1.31	0.19	270	~
24	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:BOOSTER:COUNTRY	0.89	0.13	-0.75	-1.30	0.20	270	~
16	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:FACTORS:BASE	0.89	0.12	-0.90	-1.63	0.10	270	~
21	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:BOOSTER:ALL	0.88	0.13	-1.22	-2.14	< 0.05	270	-
65	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:BOOSTER:ALL	0.88	0.13	-1.35	-1.98	< 0.05	270	-
20	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:FACTORS:LOCALITY	0.88	0.13	-1.68	-2.99	< 0.01	270	-
58	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:FACTORS:ALL	0.88	0.12	-1.74	-2.70	< 0.01	270	-
64	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:FACTORS:LOCALITY	0.88	0.13	-2.10	-2.96	< 0.01	270	-
15	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:FACTORS:IDs	0.85	0.14	-5.26	-7.14	< 0.01	270	-

Table J.4: continued on the next page

#	case	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
22	COLD	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD:BOOSTER:IDs	0.84	0.15	-6.41	-7.54	< 0.01	270	-
59	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:FACTORS:IDs	0.84	0.15	-6.52	-6.92	< 0.01	270	-
66	COLD_FILT	NDCG	COLD:OFFSET:ALL	0.90	0.12	COLD_FILT:BOOSTER:IDs	0.83	0.15	-7.35	-8.10	< 0.01	270	-
72	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:OFFSET:ALL	0.53	0.06	8.85	12.26	< 0.01	270	+
38	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:BOOSTER:COUNTRY	0.47	0.00	-4.41	-121.06	< 0.01	270	-
40	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:BOOSTER:LANGUAGE	0.47	0.00	-4.57	-132.37	< 0.01	270	-
37	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:BOOSTER:BASE	0.47	0.00	-4.60	-138.82	< 0.01	270	-
41	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:BOOSTER:LOCALITY	0.46	0.00	-4.67	-137.67	< 0.01	270	-
39	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:BOOSTER:DIMENSION	0.46	0.00	-4.82	-137.33	< 0.01	270	-
31	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:FACTORS:COUNTRY	0.44	0.00	-9.68	-188.77	< 0.01	270	-
35	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:BOOSTER:ALL	0.44	0.00	-9.72	-239.88	< 0.01	270	-
33	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:FACTORS:LANGUAGE	0.43	0.00	-11.55	-259.86	< 0.01	270	-
86	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:BOOSTER:LOCALITY	0.43	0.06	-11.61	-14.66	< 0.01	270	-
32	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:FACTORS:DIMENSION	0.43	0.00	-11.69	-264.01	< 0.01	270	-
82	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:BOOSTER:BASE	0.43	0.06	-11.81	-14.88	< 0.01	270	-
83	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:BOOSTER:COUNTRY	0.42	0.06	-13.25	-18.41	< 0.01	270	-
77	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:FACTORS:DIMENSION	0.42	0.04	-13.79	-25.73	< 0.01	270	-
85	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:BOOSTER:LANGUAGE	0.42	0.06	-13.93	-18.75	< 0.01	270	-

Table J.4: continued on the next page

#	case	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
84	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:BOOSTER:DIMENSION	0.42	0.06	-13.99	-18.93	< 0.01	270	-
75	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:FACTORS:BASE	0.42	0.04	-14.48	-25.79	< 0.01	270	-
76	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:FACTORS:COUNTRY	0.41	0.05	-16.86	-25.87	< 0.01	270	-
28	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:FACTORS:ALL	0.40	0.02	-17.09	-73.10	< 0.01	270	-
30	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:FACTORS:BASE	0.40	0.03	-17.64	-40.58	< 0.01	270	-
80	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:BOOSTER:ALL	0.38	0.07	-22.71	-25.10	< 0.01	270	-
78	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:FACTORS:LANGUAGE	0.37	0.07	-24.28	-29.61	< 0.01	270	-
73	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:FACTORS:ALL	0.34	0.06	-29.52	-37.71	< 0.01	270	-
79	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:FACTORS:LOCALITY	0.33	0.08	-32.67	-34.05	< 0.01	270	-
34	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:FACTORS:LOCALITY	0.31	0.07	-36.65	-41.51	< 0.01	270	-
29	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:FACTORS:IDs	0.16	0.00	-67.66	-1063.31	< 0.01	270	-
74	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:FACTORS:IDs	0.14	0.07	-70.87	-76.26	< 0.01	270	-
36	COLD	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD:BOOSTER:IDs	0.11	0.00	-77.69	-1754.72	< 0.01	270	-
81	COLD_FILT	R^2	COLD:OFFSET:ALL	0.49	0.00	COLD_FILT:BOOSTER:IDs	0.11	0.09	-77.82	-67.01	< 0.01	270	-

Table J.4: Offline user tests

Table J.4 above shows offline user tests using selected users sample in following situations: REG when there are user ratings in the training dataset, FILT when all ratings are pre-filtered on user inferred cultural dimension, COLD when there are no user ratings in the training dataset (cold-start), COLD_FILT when using pre-filtering with the inferred cultural dimension in the cold-start cases.

J.2.3 Recommender Performance in "Timeline" Tests

In the timeline tests, we add more ratings on a weekly basis with the aim to find out the best model and context feature set combinations regarding recommender system performance compared with the OFFSET and BOOSTER:BASE recommendation strategies. Figures depict the overall recommendation performance over time. Figure J.8 shows the performance of baseline recommendation strategies which are based on average ratings, while Figures J.9, J.10 and J.11 show the performance of the context-aware recommendation strategies considering feature sets with additional data extracted from Twitter and IMDB sources. Tables J.5 and J.6 show the t-test results of paired two-tail t-tests, in which all distributions of paired differences were tested using Shapiro test provided by the Scipy Python library.

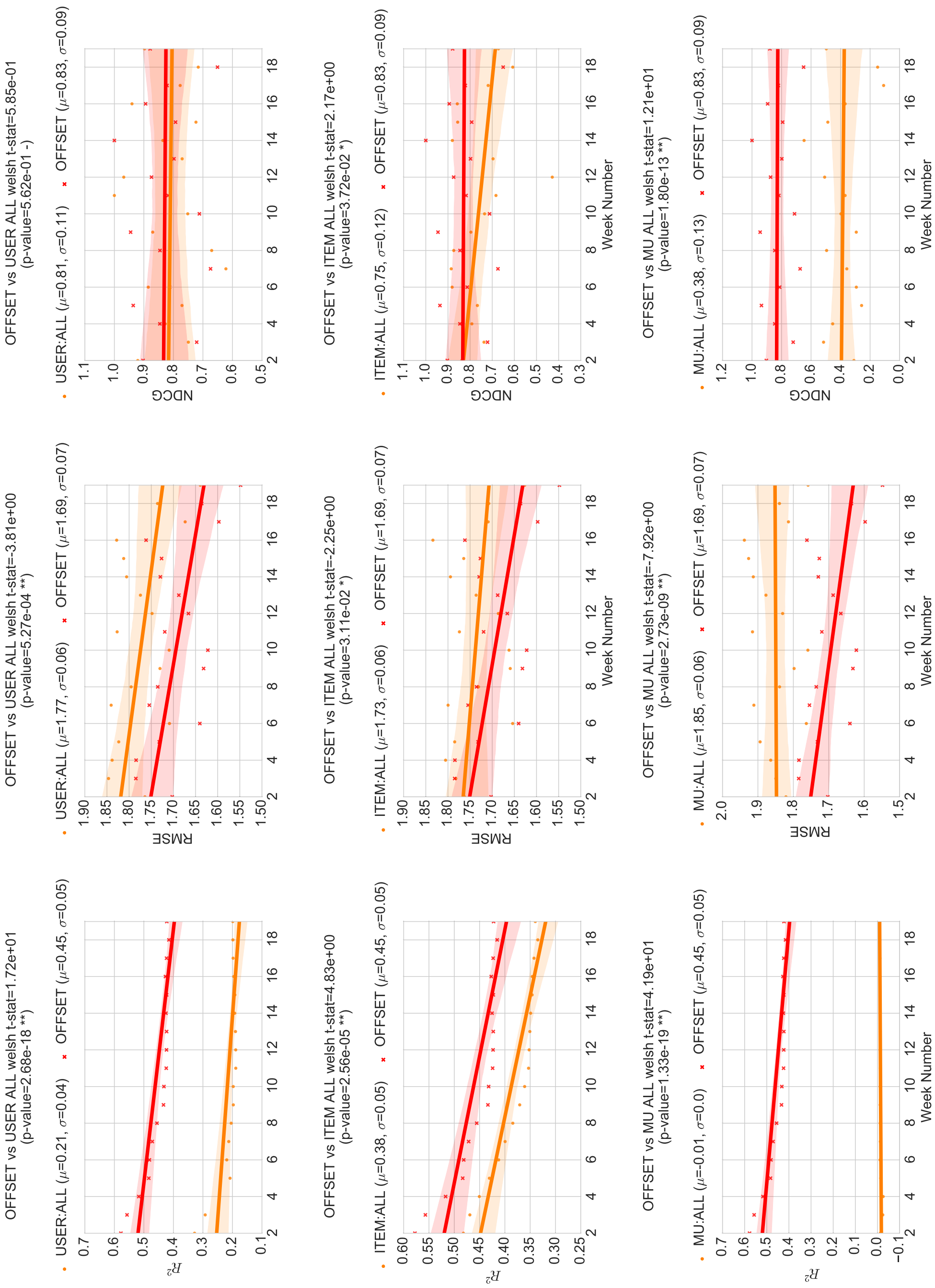
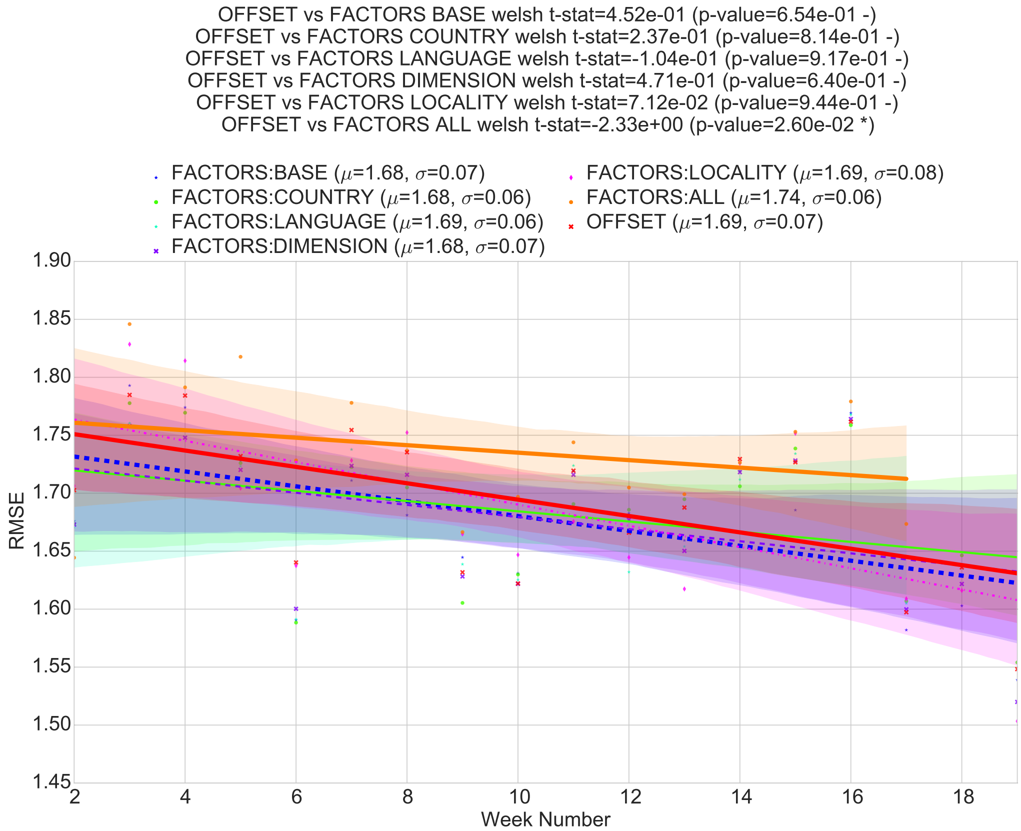


Figure J.8: Baseline strategies in timeline tests: rating prediction performance



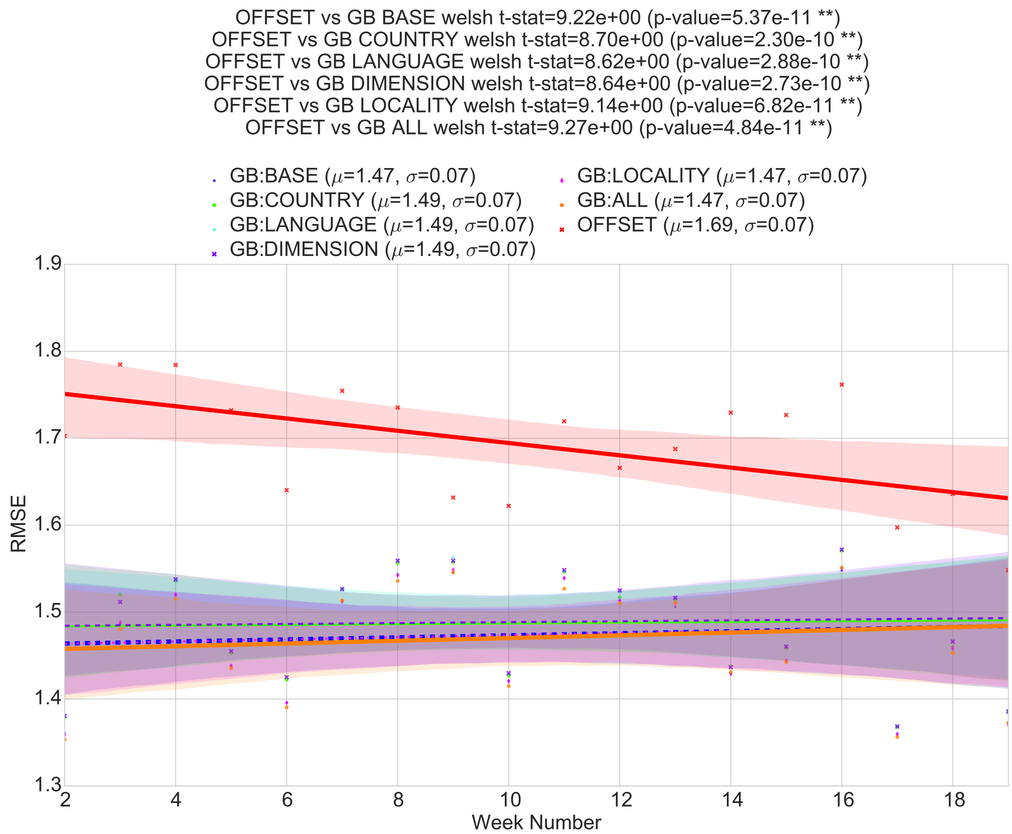
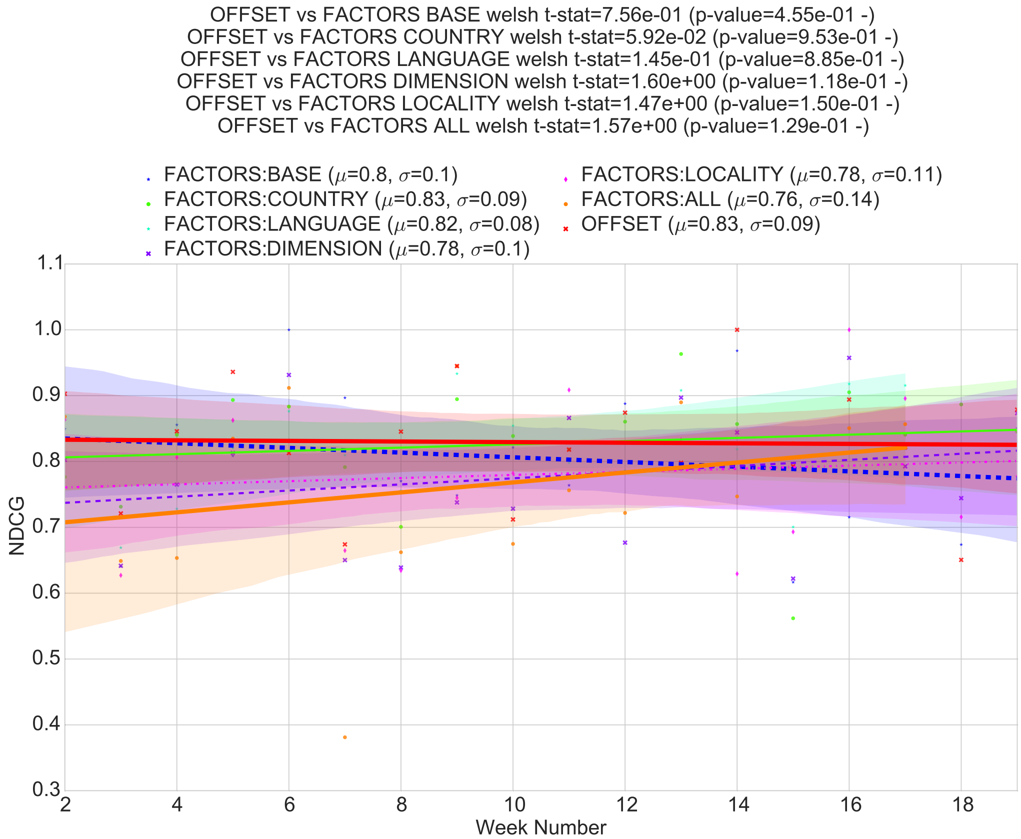


Figure J.9: Context-aware recommendation strategies in timeline: RMSE



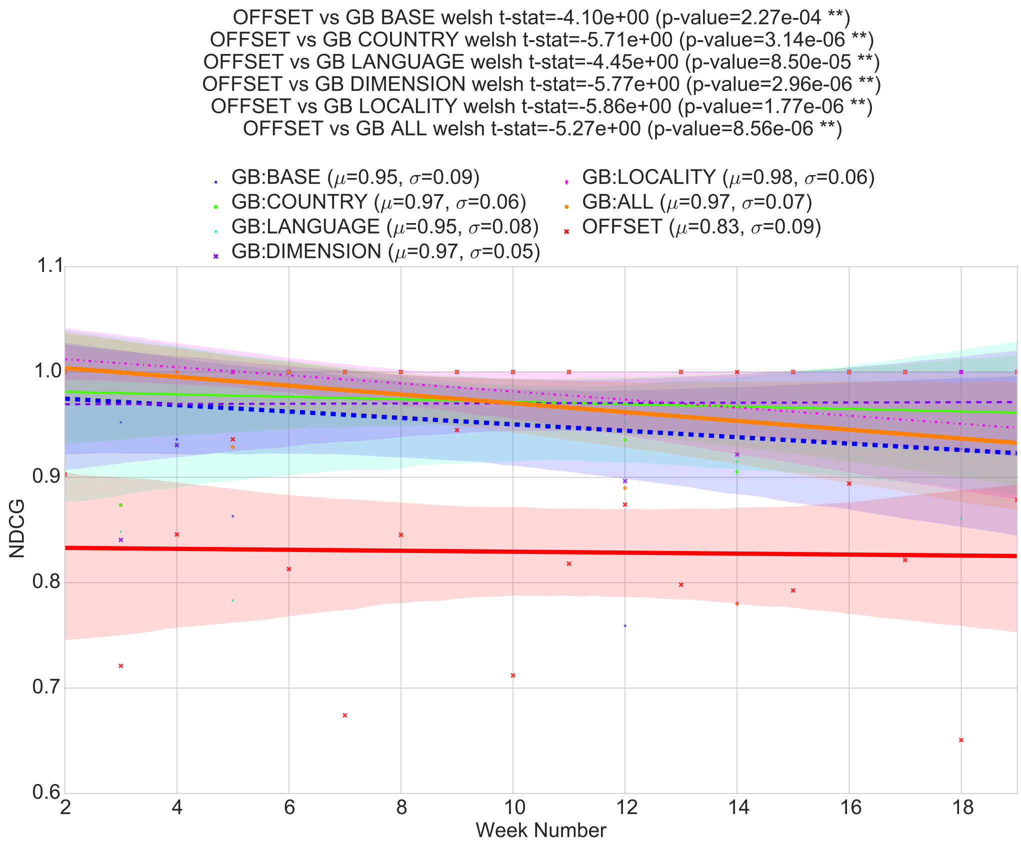
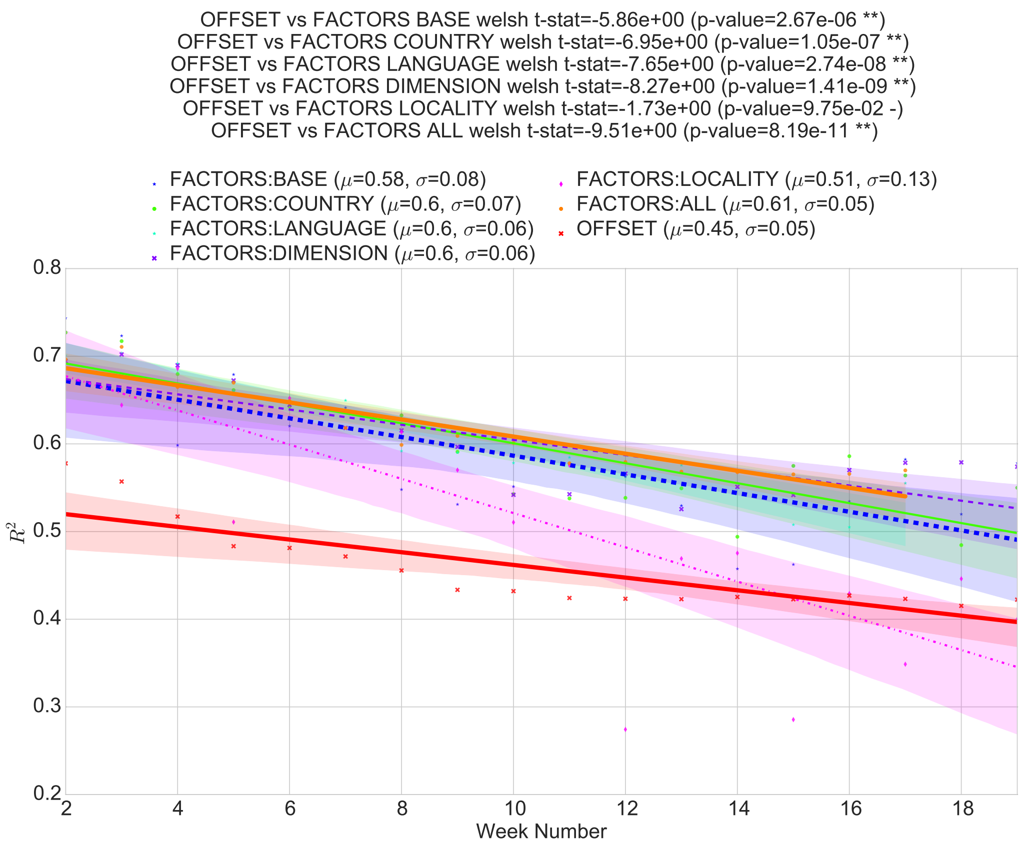


Figure J.10: Context-aware recommendation strategies in timeline: NDCG



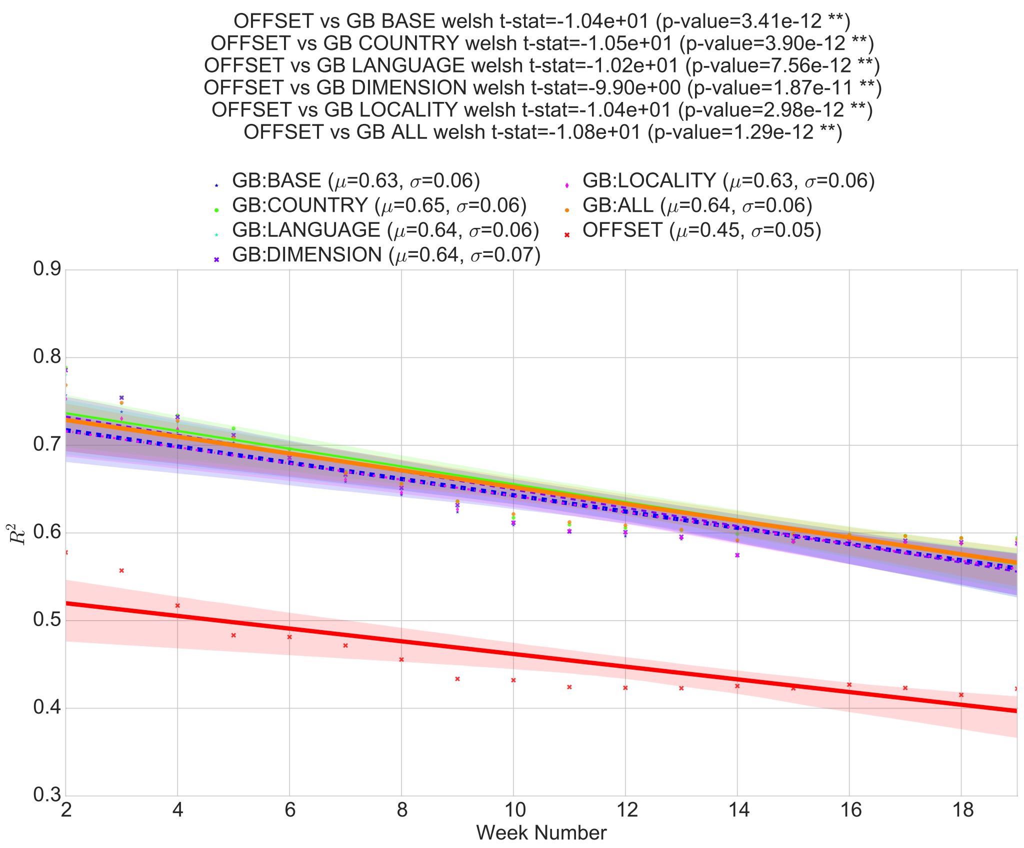


Figure J.11: Context-aware recommendation strategies in timeline: R^2

#	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	weeks	
8	RMSE	REG:OFFSET:ALL	1.69	0.07	REG:BOOSTER:IDs	1.76	0.06	4.15	8.77	< 0.01	19	-
0	RMSE	REG:OFFSET:ALL	1.70	0.06	REG:FACTORS:ALL	1.74	0.06	1.87	3.30	< 0.01	16	-
5	RMSE	REG:OFFSET:ALL	1.69	0.07	REG:FACTORS:LOCALITY	1.69	0.08	-0.39	-0.85	0.41	18	~
4	RMSE	REG:OFFSET:ALL	1.69	0.07	REG:FACTORS:COUNTRY	1.68	0.06	-0.59	-2.11	0.05	18	~
3	RMSE	REG:OFFSET:ALL	1.70	0.06	REG:FACTORS:LANGUAGE	1.69	0.06	-0.89	-3.16	< 0.01	16	+
2	RMSE	REG:OFFSET:ALL	1.69	0.07	REG:FACTORS:BASE	1.68	0.07	-0.90	-2.84	< 0.05	18	+
6	RMSE	REG:OFFSET:ALL	1.69	0.07	REG:FACTORS:DIMENSION	1.68	0.07	-0.91	-4.03	< 0.01	18	+
1	RMSE	REG:OFFSET:ALL	1.69	0.07	REG:FACTORS:IDs	1.66	0.08	-1.73	-3.33	< 0.01	18	+
10	RMSE	REG:OFFSET:ALL	1.69	0.07	REG:BOOSTER:LANGUAGE	1.49	0.07	-11.75	-10.62	< 0.01	19	+
13	RMSE	REG:OFFSET:ALL	1.69	0.07	REG:BOOSTER:DIMENSION	1.49	0.07	-11.76	-10.66	< 0.01	19	+
11	RMSE	REG:OFFSET:ALL	1.69	0.07	REG:BOOSTER:COUNTRY	1.49	0.07	-11.85	-10.84	< 0.01	19	+
12	RMSE	REG:OFFSET:ALL	1.69	0.07	REG:BOOSTER:LOCALITY	1.47	0.07	-12.60	-10.94	< 0.01	19	+
9	RMSE	REG:OFFSET:ALL	1.69	0.07	REG:BOOSTER:BASE	1.47	0.07	-12.62	-11.03	< 0.01	19	+
7	RMSE	REG:OFFSET:ALL	1.69	0.07	REG:BOOSTER:ALL	1.47	0.07	-12.78	-11.00	< 0.01	19	+
26	NDCG	REG:OFFSET:ALL	0.83	0.09	REG:BOOSTER:LOCALITY	0.98	0.06	17.94	5.03	< 0.01	19	+
25	NDCG	REG:OFFSET:ALL	0.83	0.09	REG:BOOSTER:COUNTRY	0.97	0.06	17.10	5.55	< 0.01	19	+
27	NDCG	REG:OFFSET:ALL	0.83	0.09	REG:BOOSTER:DIMENSION	0.97	0.05	17.07	5.72	< 0.01	19	+

Table J.5 continued on the next page

#	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	weeks	
21	NDCG	REG:OFFSET:ALL	0.83	0.09	REG:BOOSTER:ALL	0.97	0.07	16.55	4.56	< 0.01	19	+
24	NDCG	REG:OFFSET:ALL	0.83	0.09	REG:BOOSTER:LANGUAGE	0.95	0.08	14.75	4.45	< 0.01	19	+
23	NDCG	REG:OFFSET:ALL	0.83	0.09	REG:BOOSTER:BASE	0.95	0.09	14.27	3.57	< 0.01	19	+
18	NDCG	REG:OFFSET:ALL	0.83	0.09	REG:FACTORS:COUNTRY	0.83	0.09	-0.23	-0.07	0.95	18	~
17	NDCG	REG:OFFSET:ALL	0.84	0.09	REG:FACTORS:LANGUAGE	0.82	0.08	-1.48	-0.50	0.63	16	~
16	NDCG	REG:OFFSET:ALL	0.83	0.09	REG:FACTORS:BASE	0.80	0.10	-2.95	-0.90	0.38	18	~
19	NDCG	REG:OFFSET:ALL	0.83	0.09	REG:FACTORS:LOCALITY	0.78	0.11	-5.83	-1.65	0.12	18	~
20	NDCG	REG:OFFSET:ALL	0.83	0.09	REG:FACTORS:DIMENSION	0.78	0.10	-6.30	-2.01	0.06	18	~
22	NDCG	REG:OFFSET:ALL	0.83	0.09	REG:BOOSTER:IDs	0.76	0.12	-7.91	-2.09	0.05	19	~
14	NDCG	REG:OFFSET:ALL	0.84	0.09	REG:FACTORS:ALL	0.76	0.14	-8.69	-2.50	< 0.05	16	-
15	NDCG	REG:OFFSET:ALL	0.83	0.09	REG:FACTORS:IDs	0.75	0.20	-9.25	-1.22	0.24	18	~
39	R^2	REG:OFFSET:ALL	0.45	0.05	REG:BOOSTER:COUNTRY	0.65	0.06	41.96	44.12	< 0.01	19	+
35	R^2	REG:OFFSET:ALL	0.45	0.05	REG:BOOSTER:ALL	0.64	0.06	41.35	50.44	< 0.01	19	+
38	R^2	REG:OFFSET:ALL	0.45	0.05	REG:BOOSTER:LANGUAGE	0.64	0.06	40.56	43.85	< 0.01	19	+
41	R^2	REG:OFFSET:ALL	0.45	0.05	REG:BOOSTER:DIMENSION	0.64	0.07	40.54	38.73	< 0.01	19	+
37	R^2	REG:OFFSET:ALL	0.45	0.05	REG:BOOSTER:BASE	0.63	0.06	39.37	51.14	< 0.01	19	+
40	R^2	REG:OFFSET:ALL	0.45	0.05	REG:BOOSTER:LOCALITY	0.63	0.06	39.22	47.01	< 0.01	19	+

Table J.5 continued on the next page

#	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	weeks	
28	R^2	REG:OFFSET:ALL	0.46	0.05	REG:FACTORS:ALL	0.61	0.05	32.99	37.42	< 0.01	16	+
34	R^2	REG:OFFSET:ALL	0.46	0.05	REG:FACTORS:DIMENSION	0.60	0.06	31.50	25.59	< 0.01	18	+
31	R^2	REG:OFFSET:ALL	0.46	0.05	REG:FACTORS:LANGUAGE	0.60	0.06	31.06	19.38	< 0.01	16	+
32	R^2	REG:OFFSET:ALL	0.46	0.05	REG:FACTORS:COUNTRY	0.60	0.07	30.36	17.92	< 0.01	18	+
30	R^2	REG:OFFSET:ALL	0.46	0.05	REG:FACTORS:BASE	0.58	0.08	27.33	11.49	< 0.01	18	+
33	R^2	REG:OFFSET:ALL	0.46	0.05	REG:FACTORS:LOCALITY	0.51	0.13	12.03	2.30	< 0.05	18	+
29	R^2	REG:OFFSET:ALL	0.46	0.05	REG:FACTORS:IDs	0.11	0.02	-75.37	-20.04	< 0.01	18	-
36	R^2	REG:OFFSET:ALL	0.45	0.05	REG:BOOSTER:IDs	0.03	0.01	-92.46	-34.14	< 0.01	19	-

Table J.5: OFFSET recommender versus other in “Timeline” tests

#	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
1	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:BOOSTER:IDs	1.76	0.06	19.19	16.66	< 0.01	19	-
8	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:FACTORS:ALL	1.74	0.06	17.91	15.06	< 0.01	16	-
12	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:FACTORS:LOCALITY	1.69	0.08	14.95	11.85	< 0.01	18	-
10	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:FACTORS:LANGUAGE	1.69	0.06	14.72	13.58	< 0.01	16	-
11	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:FACTORS:COUNTRY	1.68	0.06	14.72	14.37	< 0.01	18	-
6	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:OFFSET:ALL	1.69	0.07	14.44	11.03	< 0.01	19	-
7	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:FACTORS:BASE	1.68	0.07	14.37	14.33	< 0.01	18	-
13	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:FACTORS:DIMENSION	1.68	0.07	14.35	14.57	< 0.01	18	-
9	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:FACTORS:IDs	1.66	0.08	13.41	10.25	< 0.01	18	-
2	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:BOOSTER:LANGUAGE	1.49	0.07	0.99	9.67	< 0.01	19	-
5	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:BOOSTER:DIMENSION	1.49	0.07	0.98	9.55	< 0.01	19	-
3	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:BOOSTER:COUNTRY	1.49	0.07	0.87	7.00	< 0.01	19	-
4	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:BOOSTER:LOCALITY	1.47	0.07	0.02	0.78	0.45	19	~
0	RMSE	REG:BOOSTER:BASE	1.47	0.07	REG:BOOSTER:ALL	1.47	0.07	-0.19	-2.73	< 0.05	19	+
18	NDCG	REG:BOOSTER:BASE	0.95	0.09	REG:BOOSTER:LOCALITY	0.98	0.06	3.21	2.37	< 0.05	19	+
17	NDCG	REG:BOOSTER:BASE	0.95	0.09	REG:BOOSTER:COUNTRY	0.97	0.06	2.48	1.68	0.11	19	~
19	NDCG	REG:BOOSTER:BASE	0.95	0.09	REG:BOOSTER:DIMENSION	0.97	0.05	2.45	1.62	0.12	19	~

Table J.6: continued on the next page

#	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
14	NDCG	REG:BOOSTER:BASE	0.95	0.09	REG:BOOSTER:ALL	0.97	0.07	2.00	2.32	< 0.05	19	+
16	NDCG	REG:BOOSTER:BASE	0.95	0.09	REG:BOOSTER:LANGUAGE	0.95	0.08	0.42	0.26	0.80	19	~
20	NDCG	REG:BOOSTER:BASE	0.95	0.09	REG:OFFSET:ALL	0.83	0.09	-12.49	-3.57	< 0.01	19	-
25	NDCG	REG:BOOSTER:BASE	0.96	0.08	REG:FACTORS:COUNTRY	0.83	0.09	-13.52	-4.17	< 0.01	18	-
24	NDCG	REG:BOOSTER:BASE	0.96	0.08	REG:FACTORS:LANGUAGE	0.82	0.08	-13.70	-5.20	< 0.01	16	-
21	NDCG	REG:BOOSTER:BASE	0.96	0.08	REG:FACTORS:BASE	0.80	0.10	-15.88	-4.38	< 0.01	18	-
26	NDCG	REG:BOOSTER:BASE	0.96	0.08	REG:FACTORS:LOCALITY	0.78	0.11	-18.38	-6.40	< 0.01	18	-
27	NDCG	REG:BOOSTER:BASE	0.96	0.08	REG:FACTORS:DIMENSION	0.78	0.10	-18.79	-6.14	< 0.01	18	-
15	NDCG	REG:BOOSTER:BASE	0.95	0.09	REG:BOOSTER:IDs	0.76	0.12	-19.41	-6.19	< 0.01	19	-
22	NDCG	REG:BOOSTER:BASE	0.96	0.08	REG:FACTORS:ALL	0.76	0.14	-20.02	-4.91	< 0.01	16	-
23	NDCG	REG:BOOSTER:BASE	0.96	0.08	REG:FACTORS:IDs	0.75	0.20	-21.34	-4.47	< 0.01	18	-
31	R^2	REG:BOOSTER:BASE	0.63	0.06	REG:BOOSTER:COUNTRY	0.65	0.06	1.86	8.69	< 0.01	19	+
28	R^2	REG:BOOSTER:BASE	0.63	0.06	REG:BOOSTER:ALL	0.64	0.06	1.42	11.41	< 0.01	19	+
30	R^2	REG:BOOSTER:BASE	0.63	0.06	REG:BOOSTER:LANGUAGE	0.64	0.06	0.85	4.47	< 0.01	19	+
33	R^2	REG:BOOSTER:BASE	0.63	0.06	REG:BOOSTER:DIMENSION	0.64	0.07	0.84	2.63	< 0.05	19	+
32	R^2	REG:BOOSTER:BASE	0.63	0.06	REG:BOOSTER:LOCALITY	0.63	0.06	-0.10	-0.59	0.56	19	~
36	R^2	REG:BOOSTER:BASE	0.64	0.06	REG:FACTORS:ALL	0.61	0.05	-4.54	-8.34	< 0.01	16	-

Table J.6: continued on the next page

#	metric	baseline	μ_1	σ_1	tested	μ_2	σ_2	change (%)	t	p	users	
41	R^2	REG:BOOSTER:BASE	0.64	0.06	REG:FACTORS:DIMENSION	0.60	0.06	-5.67	-8.23	< 0.01	18	-
38	R^2	REG:BOOSTER:BASE	0.64	0.06	REG:FACTORS:LANGUAGE	0.60	0.06	-5.92	-7.16	< 0.01	16	-
39	R^2	REG:BOOSTER:BASE	0.64	0.06	REG:FACTORS:COUNTRY	0.60	0.07	-6.49	-6.41	< 0.01	18	-
35	R^2	REG:BOOSTER:BASE	0.64	0.06	REG:FACTORS:BASE	0.58	0.08	-8.66	-5.40	< 0.01	18	-
40	R^2	REG:BOOSTER:BASE	0.64	0.06	REG:FACTORS:LOCALITY	0.51	0.13	-19.63	-5.64	< 0.01	18	-
34	R^2	REG:BOOSTER:BASE	0.63	0.06	REG:OFFSET:ALL	0.45	0.05	-28.25	-51.14	< 0.01	19	-
37	R^2	REG:BOOSTER:BASE	0.64	0.06	REG:FACTORS:IDs	0.11	0.02	-82.34	-27.24	< 0.01	18	-
29	R^2	REG:BOOSTER:BASE	0.63	0.06	REG:BOOSTER:IDs	0.03	0.01	-94.59	-41.21	< 0.01	19	-

Table J.6: BOOSTER:BASE recommender versus other in “Timeline” tests

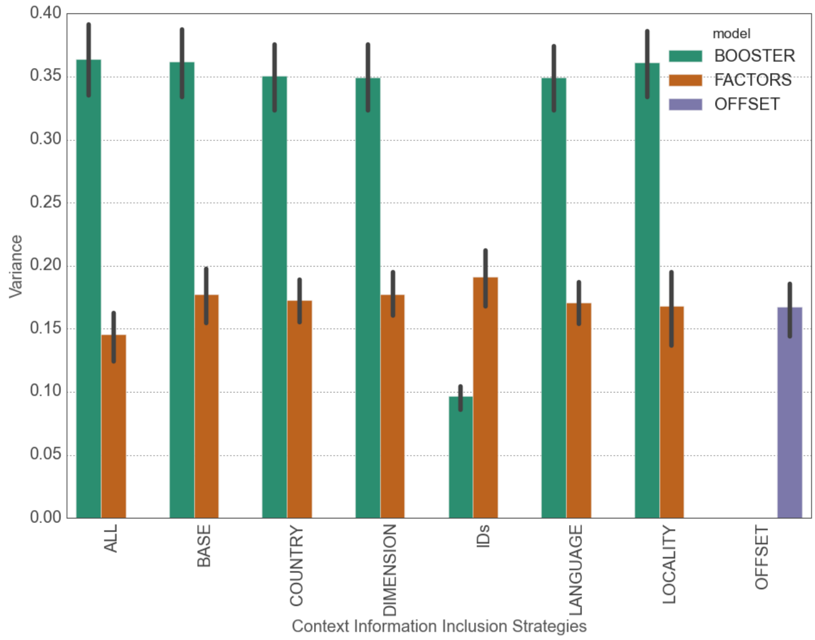


Figure J.12: Explained variance in timeline tests with different context inclusion strategies

J.2.4 Variance and Context Variables in "Timeline" Tests

Figure J.12 shows the explained variance metric performance in respect to used feature sets (we call the related parameters as context variables). We can see that the BOOSTER model outperforms OFFSET (average-based) and FACTORS models for all context inclusion strategies except "IDs", which denotes usage of only user and movie labels (binarised). Interestingly, the IDs-based strategy can provide the best performance for FACTORS model, which is, however, competing closely with the OFFSET. Even though that we might prefer the BOOSTER model, we must be aware of the relatively a good variance when using BASE strategy, considering only averages together with movie and user labels.

References

[1] F. Abel, Q. Gao, G. Houben, and K. Tao. Analyzing temporal dynamics in Twitter profiles for personalized recommendations in the social web. In *Proceedings of the 3rd International Web Science Conference*, page 2. ACM, 2011.

[2] F. Abel, Q. Gao, G. Houben, and K. Tao. Analyzing user modeling on Twitter for personalized news recommendations. *User Modeling, Adaption and Personalization*, pages 1–12, 2011.

[3] F. Abel, E. Herder, G. Houben, N. Henze, and D. Krause. Cross-system user modeling and personalization on the social web. *User Modeling and User-Adapted Interaction (UMUAI), Special Issue on Personalization in Social Web Systems*, 22(3):1–42, 2011.

[4] A. Acar. Culture and social media usage: Analysis of Japanese Twitter users. *International Journal of Electronic Commerce Studies*, 4(1):21–32, 2013.

[5] P. Adamopoulos and A. Tuzhilin. Estimating the value of multi-dimensional data sets in context-based recommender systems, 2014.

[6] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749, 2005.

[7] G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. In *Recommender systems handbook*, pages 217–253. Springer, 2011.

[8] D. Agarwal, B.-C. Chen, and B. Long. Localized factor models for multi-context recommendation. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 609–617. ACM, 2011.

[9] I. Ahmad. Global internet, mobile and social media engagement and usage stats and facts [infographic], 2013.

[10] A. S. Al-Harthi. Distance higher education experiences of Arab Gulf students in the United States: A cultural perspective. *The International Review of Research in Open and Distributed Learning*, 6(3), 2006.

[11] F. Al Zamal, W. Liu, and D. Ruths. Homophily and latent attribute inference: Inferring latent attributes of Twitter users from neighbors. In *ICWSM*, 2012.

[12] M.-d.-C. Alarcon-del Amo, M.-A. Gomez-Borja, and C. Lorenzo-Romero. Are the users of social networking sites homogeneous? a cross-cultural study. *Frontiers in psychology*, 6:1127, 2015.

[13] B. Alex, C. Llewellyn, C. Grover, J. Oberlander, and R. Tobin. Homing in on Twitter users: Evaluating an enhanced geoparser for user profile locations. In *LREC*, 2016.

[14] R. Alexander, D. Murray, and N. Thompson. Cross-cultural Web usability model. In *International Conference on Web Information Systems Engineering*, pages 75–89. Springer, 2017.

[15] J. Allik and R. McCrae. Toward a geography of personality traits. *Journal of Cross-Cultural Psychology*, 35(1):13–28, 2004.

[16] D. G. Altman and J. M. Bland. Statistics notes: the normal distribution. *Bmj*, 310(6975):298, 1995.

[17] Y. Artzi, P. Pantel, and M. Gamon. Predicting responses to microblog posts. In *Proceedings of the 2012 Conference of the North American Chapter of the*

Association for Computational Linguistics: Human Language Technologies, pages 602–606. Association for Computational Linguistics, 2012.

[18] C. A. Bail. The cultural environment: Measuring culture with big data. *Theory and Society*, 43(3-4):465–482, 2014.

[19] A. Bassolas, M. Lenormand, A. Tugores, B. Gonçalves, and J. J. Ramasco. Touristic site attractiveness seen through Twitter. *EPJ Data Science*, 5(1):12, 2016.

[20] R. J. Bayardo and R. Agrawal. Data privacy through optimal k-anonymization. In *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on*, pages 217–228. IEEE, 2005.

[21] T. Berners-Lee. Universal resource identifiers in WWW: a unifying syntax for the expression of names and addresses of objects on the network as used in the World-wide Web (no. rfc 1630). `https://www.rfc-editor.org/rfc/pdfrfc/rfc1630.txt.pdf`, 1994.

[22] T. Berners-Lee, W. Hall, J. A. Hendler, K. O'Hara, N. Shadbolt, and D. J. Weitzner. A framework for Web science. *Foundations and trends in Web Science*, 1(1):1–130, 2006.

[23] M. C. Bettoni and C. Eggs. User-centred knowledge management: A constructivist and socialized view. *Constructivist Foundations*, 5(3), 2010.

[24] S. Beugelsdijk, R. Maseland, and A. Van Hoorn. Are Hofstede's culture dimensions stable over time? a generational cohort analysis. *A Generational Cohort Analysis (October 7, 2013)*, 2013.

[25] M. Bieliková, M. Kompan, and D. Zeleník. Effective hierarchical vector-based news representation for personalized recommendation. *Computer Science and Information Systems*, 9(1):303–322, 2012.

[26] D. Boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pages 1–10. IEEE, 2010.

[27] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.

[28] P. Brusilovsky. Adaptive hypermedia: An attempt to analyze and generalize. *Multimedia, Hypermedia, and Virtual Reality Models, Systems, and Applications*, pages 288–304, 1996.

[29] P. Brusilovsky and E. Millán. User models for adaptive hypermedia and adaptive educational systems. *The adaptive web*, pages 3–53, 2007.

[30] M. Bulearca and S. Bulearca. Twitter: a viable marketing tool for SMEs. *Global Business and Management Research: An International Journal*, 2(4):296–309, 2010.

[31] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning*, pages 89–96. ACM, 2005.

[32] R. Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370, 2002.

[33] Y. Cai, H.-f. Leung, Q. Li, H. Min, J. Tang, and J. Li. Typicality-based collaborative filtering recommendation. *IEEE Trans. Knowl. Data Eng.*, 26(3):766–779, 2014.

[34] A. Caliskan Islam, J. Walsh, and R. Greenstadt. Privacy detective: Detecting private information and collective privacy behavior in a large social network. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, pages 35–46. ACM, 2014.

[35] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi. Measuring user influence in Twitter: The million follower fallacy. In *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2010.

[36] M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the Flickr social network. In *Proceedings of the 18th international conference on World wide web*, pages 721–730. ACM, 2009.

[37] Y. Chan and R. P. Walmsley. Learning and understanding the Kruskal-Wallis one-way analysis-of-variance-by-ranks test for differences among three or more independent groups. *Physical therapy*, 77(12):1755–1761, 1997.

[38] J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy. Make new friends, but keep the old: recommending people on social networking sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 201–210. ACM, 2009.

[39] Y. Chen, J. Zhao, X. Hu, X. Zhang, Z. Li, and T.-S. Chua. From interest to function: Location estimation in social media. In *AAAI*, 2013.

[40] Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: a content-based approach to geo-locating Twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 759–768. ACM, 2010.

[41] M. J. Chorley, R. M. Whitaker, and S. M. Allen. Personality and location-based social networks. *Computers in Human Behavior*, 46:45–56, 2015.

[42] J. Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.

[43] R. Compton, M. S. Keegan, and J. Xu. Inferring the geographic focus of online documents from social media sharing patterns. *arXiv preprint arXiv:1406.2392*, 2014.

[44] R. Cooley, B. Mobasher, and J. Srivastava. Web mining: Information and pattern discovery on the World wide web. In *Tools with Artificial Intelligence, 1997. Proceedings., Ninth IEEE International Conference on*, pages 558–567. IEEE, 1997.

[45] crossculture.com. The Lewis model dimensions of behaviour. `https://www.crossculture.com/the-lewis-model-dimensions-of-behaviour/`, 2015.

[46] crunchbase.com. Cambridge Analytica scandal: the biggest revelations so far. `https://www.theguardian.com/uk-news/2018/mar/22/cambridge-analytica-scandal-the-biggest-revelations-so-far`, 2018.

[47] crunchbase.com. Cask: Cask provides the first unified integration platform for Hadoop and Apache Spark. `https://www.crunchbase.com/organization/cask`, 2018.

[48] crunchbase.com. Startup battlefield Europe 2018 winner: Wingly. `https://techcrunch.com/video-article/wingly-presents-at-startup-battlefield-europe-finals/`, 2018.

[49] crunchbase.com. Wingly. `https://www.crunchbase.com/organization/wingly`, 2018.

[50] A. C. Curry and V. Rieser. # MeToo Alexa: How conversational systems respond to sexual harassment. In *Proceedings of the Second ACL Workshop on Ethics in Natural Language Processing*, pages 7–14, 2018.

[51] E. Daehnhardt, , N. K. Taylor, and Y. Jing. Mining microblogs to exploit culture-awareness in Web adaptation. In *SICSA PhD Conference*. University of Glasgow, 2015.

[52] E. Daehnhardt, Y. Jing, and N. Taylor. Cultural and geolocation aspects of communication in Twitter. In *Social Informatics, 2014 ASE/IEEE International Conference on*, pages 1–12. ASE, Academy of Science and Engineering, USA, 2014.

[53] E. Daehnhardt, N. K. Taylor, and Y. Jing. Usage and consequences of privacy settings in microblogs. In *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on*, pages 667–674. IEEE, 2015.

[54] R. D'AGOSTINO and E. S. Pearson. Tests for departure from normality. empirical results for the distributions of b 2 and b. *Biometrika*, 60(3):613–622, 1973.

[55] R. B. d'Agostino. An omnibus test of normality for moderate and large size samples. *Biometrika*, 58(2):341–348, 1971.

[56] H. Dai and P. C. Palvi. Mobile commerce adoption in China and the United States: a cross-cultural study. *ACM SIGMIS Database*, 40(4):43–61, 2009.

[57] E. De Cristofaro, C. Soriente, G. Tsudik, and A. Williams. Hummingbird: Privacy at the time of Twitter. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 285–299. IEEE, 2012.

[58] A. Dean, D. Voss, D. Draguljić, et al. *Design and analysis of experiments*, volume 1. Springer, 1999.

[59] W. Dong, M. Qiu, and F. Zhu. Who am i on Twitter?: A cross-country comparison. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 253–254. International World Wide Web Conferences Steering Committee, 2014.

[60] S. Dooms, T. De Pessemier, and L. Martens. Movietweetings: a movie rating dataset collected from Twitter. In *Workshop on Crowdsourcing and Human Computation for Recommender Systems, CrowdRec at RecSys 2013*, volume 2013, 2013.

[61] P. Dourish. What we talk about when we talk about context. *Personal and ubiquitous computing*, 8(1):19–30, 2004.

[62] M. Eirinaki and M. Vazirgiannis. Web mining for Web personalization. *ACM Transactions on Internet Technology (TOIT)*, 3(1):1–27, 2003.

[63] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan. Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction*, 4(2):81–173, 2011.

[64] G. Ericson. How to choose Azure Machine Learning algorithms for clustering, classification, or regression. `https://azure.microsoft.com/en-us/documentation/articles/machine-learning-algorithm-choice/`, 2015.

[65] European Commission. 2018 reform of EU data protection rules. `https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en`, 2018.

[66] European Commission. A new era for data protection in the EU: What changes after may 2018. `https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf`, 2018.

[67] T. Fawcett. ROC graphs: Notes and practical considerations for researchers. *Machine learning*, 31:1–38, 2004.

[68] B. J. Feir-Walsh and L. E. Toothaker. An empirical comparison of the ANOVA F-test, normal scores test and Kruskal-Wallis test under violation of assumptions. *Educational and Psychological Measurement*, 34(4):789–799, 1974.

[69] B. Ferwerda, M. P. Graus, A. Vall, M. Tkalcic, and M. Schedl. The influence of users' personality traits on satisfaction and attractiveness of diversified recommendation lists. In *EMPIRE@ RecSys*, pages 43–47, 2016.

[70] B. Ferwerda and M. Schedl. Investigating the relationship between diversity in music consumption behavior and cultural dimensions: A cross-country analysis. In *Proc. of the 1st Workshop on SOAP*, 2016.

[71] R. T. Fielding and R. N. Taylor. *Architectural styles and the design of network-based software architectures*. University of California, Irvine Doctoral dissertation, 2000.

[72] G. Fischer. User modeling in human–computer interaction. *User modeling and user-adapted interaction*, 11(1):65–86, 2001.

[73] J. Fleiss. The measurement of interrater agreement. statistical methods for rates and proportions, 2nd edn pp 212–304, 1981.

[74] J. L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378, 1971.

[75] K. Forbes-Riley and D. Litman. Adapting to student uncertainty improves tutoring dialogues. In *Proc. Intl. Conf. on Artificial Intelligence in Education*, 2009.

[76] K. Fujii, H. Nanba, T. Takezawa, A. Ishino, M. Okumura, and Y. Kurata. Travellers' behaviour analysis based on automatically identified attributes from travel blog entries. In *Proceedings of Workshop of Artificial Intelligence for Tourism, PRICAI*, 2016.

[77] S. Gallacher, E. Papadopoulou, N. K. Taylor, and M. H. Williams. Learning user preferences for adaptive pervasive environments: an incremental and temporal approach. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 8(1):5, 2013.

[78] Q. Gao, F. Abel, G.-J. Houben, and Y. Yu. A comparative study of users' microblogging behavior on Sina Weibo and Twitter. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 88–101. Springer, 2012.

[79] R. García-Gavilanes, Y. Mejova, and D. Quercia. Twitter ain't without frontiers: economic, social, and cultural boundaries in international communication. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 1511–1522. ACM, 2014.

[80] R. Garcia-Gavilanes, D. Quercia, and A. Jaimes. Cultural dimensions in Twitter: Time, individualism and power. *AAAI ICWSM*, 2013.

[81] R. O. Garcia Gavilanes. On the quest of discovering cultural trails in social media. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 747–752. ACM, 2013.

[82] M. Gaved, P. Luley, S. Efremidis, I. Georgiou, A. Kukulska-Hulme, A. Jones, and E. Scanlon. Challenges in context-aware mobile language learning: the MASELTOV approach. In *Mobile as a Mainstream–Towards Future Challenges in Mobile Learning*, pages 351–364. Springer, 2014.

[83] R. G. Gavilanes, N. OHare, L. M. Aiello, and A. Jaimes. Follow my friends this friday! an analysis of human-generated friendship recommendations. In *Social Informatics*, pages 46–59. Springer, 2013.

[84] C. Geertz and M. Banton. Religion as a cultural system, 1966.

[85] L. Ghahremanlou, W. Sherchan, and J. A. Thom. Geotagging Twitter messages in crisis management. *The Computer Journal*, page bxu034, 2014.

[86] A. Ghasemi and S. Zahediasl. Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, 10(2):486, 2012.

[87] M. Giatsoglou, D. Chatzakou, and A. Vakali. Community detection in social media by leveraging interactions and intensities. In *Web Information Systems Engineering–WISE 2013*, pages 57–72. Springer, 2013.

[88] D. H.-L. Goh and A. Y. Chua. Understanding the barriers to using microblogs. In *Proceedings of the World Congress on Engineering and Computer Science*, volume 1, 2013.

[89] K. Grabczewski and N. Jankowski. Feature selection with decision tree criterion. In *null*, pages 212–217. IEEE, 2005.

[90] M. Graham, S. A. Hale, and D. Gaffney. Where in the World are you? geolocation and language identification in Twitter. *The Professional Geographer*, 66(4):568–578, 2014.

[91] T. Grill. Python implementation of Krippendorff's alpha – inter-rater reliability. `https://raw.githubusercontent.com/grrrr/krippendorff-alpha/master/krippendorff_alpha.py`, 2018.

[92] J. Grimmelmann. Saving Facebook. *Iowa L. Rev.*, 94:1137, 2008.

[93] T. Gruber. Collective knowledge systems: Where the social web meets the semantic web. *Web semantics: science, services and agents on the World Wide Web*, 6(1):4–13, 2008.

[94] A. Gunawardana and G. Shani. A survey of accuracy evaluation metrics of recommendation tasks. *The Journal of Machine Learning Research*, 10:2935–2962, 2009.

[95] V. Gurusamy, S. Kannan, and J. R. Prabhu. Mining the attitude of social network users using K-means clustering. *International Journal*, 7(5), 2017.

[96] C. E. Gutierrez, M. R. Alsharif, and K. Yamashita. Uncover trending topics on data stream by linear prediction modeling. *IJCSNS*, 14(5):1, 2014.

[97] E. T. Hall et al. *The silent language*, volume 3. Doubleday New York, 1959.

[98] J. Hamill. You'll soon be able to see yourself in Facebook photos you're not tagged in using facial recognition. `https://metro.co.uk/2018/03/09/soon-able-see-photos-not-tagged-facebook-using-facial-recognition-7373845/`, 2018.

[99] K. Hampton, L. S. Goulet, L. Rainie, and K. Purcell. Social networking sites and our lives. *Retrieved July 12, 2011 from*, 2011.

[100] B. Han, P. Cook, and T. Baldwin. Text-based Twitter user geolocation prediction. *J. Artif. Intell. Res.(JAIR)*, 49:451–500, 2014.

[101] J. Hannon, M. Bennett, and B. Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 199–206. ACM, 2010.

[102] J. Hannon, E. Knutov, P. De Bra, M. Pechenizkiy, K. McCarthy, and B. Smyth. Bridging recommendation and adaptation: Generic adaptation framework-Twittomender compliance case-study. In *Second international Workshop on Dynamic and Adaptive Hypertext*, page 1, 2011.

[103] J. Hare, J. Davies, S. Samangooei, and P. Lewis. Placing photos with a multi-modal probability density function. In *Proceedings of International Conference on Multimedia Retrieval*, page 329. ACM, 2014.

[104] B. Hecht, L. Hong, B. Suh, and E. H. Chi. Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 237–246. ACM, 2011.

[105] D. Heckmann. *Ubiquitous user modeling*. PhD thesis, The Saarland University, 2006.

[106] R. Heimgärtner. Using converging strategies to reduce divergence in intercultural user interface design. *Journal of Computer and Communications*, 5(04):84, 2017.

[107] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):5–53, 2004.

[108] M. Hildebrandt, K. O'Hara, and M. Waidner. *Digital enlightenment yearbook 2013: The value of personal data*. IOS Press, 2013.

[109] G. Hofstede. The cultural relativity of organizational practices and theories. *Journal of international business studies*, 14(2):75–89, 1983.

[110] G. Hofstede. National cultures and corporate cultures. *Communication Between Cultures. Belmont, CA: Wadsworth*, 1984.

[111] G. Hofstede. A European in Asia. *Asian Journal of Social Psychology*, 10(1):16–21, 2007.

[112] G. Hofstede. Dimensionalizing cultures: The Hofstede model in context. *Online readings in psychology and culture*, 2(1):8, 2011.

[113] D. Hong, D. K. Chiu, and V. Y. Shen. Requirements elicitation for the design of context-aware applications in a ubiquitous environment. In *Proceedings of the 7th international conference on Electronic commerce*, pages 590–596. ACM, 2005.

[114] J. Huang, K. Thornton, and E. Efthimiadis. Conversational tagging in Twitter. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, pages 173–178. ACM, 2010.

[115] C. J. Huberty and J. D. Morris. Multivariate analysis versus multiple univariate analyses. *Psychological bulletin*, 105(2):302, 1989.

[116] J. Hughes. Twitter: Why the different types of retweet? `https://glassmountains.co.uk/campfire/2013/02/23/twitter-types-of-retweet/`, 2013.

[117] D. Hull. Using statistical testing in the evaluation of retrieval experiments. In *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 329–338. ACM, 1993.

[118] IBM. IBM terminology. `https://www-01.ibm.com/software/globalization/terminology/`, 2017.

[119] E. Ilina, F. Abel, and G. Houben. Mining Twitter for cultural patterns. In H. Reiterer and O. Deussen, editors, *ABIS 2012 Workshop on Personalization and Recommendation on the Web and Beyond*, pages 83–90, Mnchen, 2012. Oldenbourg Verlag.

[120] E. Ilina, C. Hauff, I. Celik, F. Abel, and G. Houben. Social event detection on Twitter. *Web Engineering*, pages 169–176, 2012.

[121] E. Ilina (Daehnhardt). A user modeling oriented analysis of cultural backgrounds in microblogging. *Human Journal*, 1(4):166–181, 2012.

[122] internetlivestats.com. 8051 tweets sent in 1 second. `http://www.internetlivestats.com/one-second/#tweets-band`.

[123] internetlivestats.com. Twitter usage statistics. `http://www.internetlivestats.com/twitter-statistics/#trend`.

[124] internetworldstats.com. Internet growth statistics: Today's road to e-commerce and global trade Internet technology reports. `https://www.internetworldstats.com/emarketing.htm`, 2018.

[125] N. Ireson and F. Ciravegna. Toponym resolution in social media. In *The Semantic Web–ISWC 2010*, pages 370–385. Springer, 2010.

[126] A. Java, X. Song, T. Finin, and B. Tseng. Why we Twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65. ACM, 2007.

[127] Y. Ji, H. Hwangbo, J. Yi, P. Rau, X. Fang, and C. Ling. The influence of cultural differences on the use of social network services and the formation of social capital. *Intl. Journal of Human–Computer Interaction*, 26(11-12):1100–1121, 2010.

[128] A. Johnson and N. Taatgen. User modeling. *The handbook of human factors in web design*, pages 424–438, 2005.

[129] D. Jurgens. That's what friends are for: Inferring location in online social media platforms based on social relationships. In *ICWSM*, 2013.

[130] A. M. Kaplan and M. Haenlein. Users of the world, unite! the challenges and opportunities of social media. *Business horizons*, 53(1):59–68, 2010.

[131] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 79–86. ACM, 2010.

[132] M. Keil, B. C. Tan, K.-K. Wei, T. Saarinen, V. Tuunainen, and A. Wassenaar. A cross-cultural study on escalation of commitment behavior in software projects. *MIS quarterly*, pages 299–325, 2000.

[133] S. King. *Hearts in Atlantis*. Simon and Schuster, 1999.

[134] R. Kling. Learning about information technologies and social change: The contribution of social informatics. *The information society*, 16(3):217–232, 2000.

[135] A. Kobsa. Generic user modeling systems. *User modeling and user-adapted interaction*, 11(1):49–63, 2001.

[136] Y. Koren. The bellkor solution to the netflix grand prize. *Netflix prize documentation*, 81:1–10, 2009.

[137] R. Kosala and H. Blockeel. Web mining research: A survey. *ACM Sigkdd Explorations Newsletter*, 2(1):1–15, 2000.

[138] H. Kosec. Pinterest vs. Instagram vs. Snapchat. `https://tandsgo.com/2017/07/pinterest-instagram-snapchat/`, 2017.

[139] A. B. Kouki. *Recommender System Performance Evaluation and Prediction: An Information Retrieval Perspective*. PhD thesis, Universidad Autónoma de Madrid, 2012.

[140] K. Krippendorff. Reliability in content analysis. *Human communication research*, 30(3):411–433, 2004.

[141] K. Krippendorff. Computing Krippendorff's alpha-reliability. `https://repository.upenn.edu/asc_papers/43`, 2011.

[142] J. Kulshrestha, F. Kooti, A. Nikravesh, and P. K. Gummadi. Geographic dissection of the Twitter network. In *ICWSM*, 2012.

[143] R. Kumar, B. Verma, and S. S. Rastogi. Social popularity based svd++ recommender system. *International Journal of Computer Applications*, 87(14):33–37, 2014.

[144] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM, 2010.

[145] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[146] B. M. Leiner, V. G. Cerf, D. D. Clark, R. E. Kahn, L. Kleinrock, D. C. Lynch, J. Postel, L. G. Roberts, and S. Wolff. A brief history of the Internet. *ACM SIGCOMM Computer Communication Review*, 39(5):22–31, 2009.

[147] X. Y. Leung, B. Bai, and K. A. Stahura. The marketing effectiveness of social media in the hotel industry a comparison of Facebook and Twitter. *Journal of Hospitality & Tourism Research*, 39(2):147–169, 2015.

[148] R. Lewis. *When cultures collide: Managing successfully across cultures.* Nicholas Brealey Publishing, 2000.

[149] R. D. Lewis. *The cultural imperative: Global trends in the 21st century.* Warnford, Hampshire S032 3LH, 2007.

[150] W. Li, P. Serdyukov, A. P. de Vries, C. Eickhoff, and M. Larson. The where in the tweet. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 2473–2476. ACM, 2011.

[151] Y. Liao, M. Moshtaghi, B. Han, S. Karunasekera, R. Kotagiri, T. Baldwin, A. Harwood, and P. Pattison. Mining micro-blogs: opportunities and challenges. In *Computational Social Networks*, pages 129–159. Springer, 2012.

[152] G. Linden, B. Smith, and J. York. Amazon. com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80, 2003.

[153] J. Lingad, S. Karimi, and J. Yin. Location extraction from disaster-related microblogs. In *Proceedings of the 22nd international conference on World Wide Web companion*, pages 1017–1020. International World Wide Web Conferences Steering Committee, 2013.

[154] M. Lui and T. Baldwin. langid.py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 system demonstrations*, pages 25–30. Association for Computational Linguistics, 2012.

[155] G. MacMillan. How Twitter users tweet while on holiday - infographic. `https://blog.twitter.com/2013/how-twitter-users-tweet-while-on-holiday-infographic`.

[156] J. Mahmud, J. Nichols, and C. Drews. Home location identification of Twitter users. *ACM Transactions on Intelligent Systems and Technology*, 5(3):47:1–47:21, 2014.

[157] H. Mao, X. Shuai, and A. Kapadia. Loose tweets: an analysis of privacy leaks on Twitter. In *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society*, pages 1–12. ACM, 2011.

[158] A. Marcus and E. W. Gould. Crosscurrents: cultural dimensions and global Web user-interface design. *interactions*, 7(4):32–46, 2000.

[159] E. McCallister, T. Grance, and K. Scarfone. Guide to protecting the confidentiality of personally identifiable information (pii): Recommendations of the national institute of standards and technology. `http://csrc.nist.gov/publications/nistpubs/800-122/sp800-122.pdf`, 2010.

[160] R. McCrae. Trait psychology and culture: Exploring intercultural comparisons. *Journal of personality*, 69(6):819–846, 2001.

[161] R. R. McCrae and A. Terracciano. Personality profiles of cultures: aggregate personality traits. *Journal of personality and social psychology*, 89(3):407, 2005.

[162] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

[163] B. McSweeney. Hofstede's model of national cultural differences and their consequences: A triumph of faith-a failure of analysis. *Human relations*, 55(1):89–118, 2002.

[164] B. Meeder, J. Tam, P. G. Kelley, and L. F. Cranor. RT@ IWantPrivacy: Widespread violation of privacy settings in the Twitter social network. In *Proceedings of the Web*, volume 2, 2010.

[165] mindtools.com. The seven dimensions of culture. `https://www.mindtools.com/pages/article/seven-dimensions.htm`, 2008.

[166] mindtools.com. The seven dimensions of culture: Your 10-minute guide to understanding cultural differences. `https://www.mindtools.com/blog/corporate/wp-content/uploads/sites/2/2014/09/Seven-Dimensions-Culture.pdf`, 2014.

[167] mith.io. Mithril: The future of social networks. `https://mith.io/whitepaper.pdf`, 2018.

[168] B. Mobasher. Data mining for Web personalization. In *The adaptive web*, pages 90–135. Springer, 2007.

[169] Y. Mohamad and C. Kouroupetroglou. User modeling. `https://www.w3.org/WAI/RD/wiki/User_modeling`, 2013.

[170] R. J. Mooney and L. Roy. Content-based book recommending using learning for text categorization. In *Proceedings of the fifth ACM conference on Digital libraries*, pages 195–204. ACM, 2000.

[171] D. Nadeau and S. Sekine. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26, 2007.

[172] H. Nakasaki, M. Kawaba, T. Utsuro, and T. Fukuhara. Mining cross-lingual/cross-cultural differences in concerns and opinions in blogs. *Computer Processing of Oriental Languages. Language Technology for the Knowledge-based Economy*, pages 213–224, 2009.

[173] C. Newberry. 28 Twitter statistics all marketers need to know in 2018. `https://blog.hootsuite.com/twitter-statistics/`, 2018.

[174] C. Newton. Twitter redesigns replies so usernames don't count against the 140-character limit. `https://www.theverge.com/2017/3/30/15115290/twitter-replies-redesign-character-limit`, 2017.

[175] T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, and J. A. Konstan. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*, pages 677–686. ACM, 2014.

[176] N. Noë, R. M. Whitaker, and S. M. Allen. Personality homophily and the local network characteristics of Facebook. In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*, pages 386–393. IEEE, 2016.

[177] J. Nothman, N. Ringland, W. Radford, T. Murphy, and J. R. Curran. Learning multilingual named entity recognition from Wikipedia. *Artificial Intelligence*, 194:151–175, 2013.

[178] oecd.org. OECD guidelines governing the protection of privacy and transborder flows of personal data. `http://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf`, 2013.

[179] O. Oh, M. Agrawal, and H. R. Rao. Information control and terrorism: Tracking the Mumbai terrorist attack through twitter. *Information Systems Frontiers*, 13(1):33–43, 2011.

[180] B. A. Olaniran. Culture, learning styles, and web 2.0. *Interactive Learning Environments*, 17(4):261–271, 2009.

[181] G. M. Olson and J. S. Olson. Distance matters. *Human-computer interaction*, 15(2):139–178, 2000.

[182] Oxford University Press. Definition of tweet in English. `https://en.oxforddictionaries.com/definition/tweet`, 2018.

[183] B. Pang, L. Lee, et al. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135, 2008.

[184] U. Panniello and M. Gorgoglione. Incorporating context into recommender systems: an empirical comparison of context-based approaches. *Electronic Commerce Research*, 12(1):1–30, 2012.

[185] U. Panniello, A. Tuzhilin, M. Gorgoglione, C. Palmisano, and A. Pedone. Experimental comparison of pre-vs. post-filtering approaches in context-aware recommender systems. In *Proceedings of the third ACM conference on Recommender systems*, pages 265–268. ACM, 2009.

[186] E. Papadopoulou, S. Gallacher, N. K. Taylor, M. H. Williams, F. R. Blackmun, I. S. Ibrahim, M. Y. Lim, I. Mimtsoudis, P. Skillen, and S. Whyte. Combining pervasive computing with social networking for a student environment. In *Proceedings of the Twelfth Australasian Symposium on Parallel and Distributed Computing-Volume 152*, pages 11–19. Australian Computer Society, Inc., 2014.

[187] E. Papadopoulou, A. Stobart, N. Taylor, and H. Williams. Enabling data subjects to remain data owners. In *Agent and Multi-Agent Systems - Technology and Applications*, pages xx–xx. ACM, Berlin: Springer Verlag, 2015.

[188] R. Pardo, M. Balliu, and G. Schneider. Formalising privacy policies in social networks. *Journal of Logical and Algebraic Methods in Programming*, 2017.

[189] E. Pariser. *The filter bubble: What the Internet is hiding from you.* Penguin UK, 2011.

[190] P. Parrish and J. Linder-VanBerschot. Cultural dimensions of learning: Addressing the challenges of multicultural instruction. *The International Review of Research in Open and Distance Learning*, 11(2):1–19, 2010.

[191] K. Pearson. *The grammar of science*, volume 20. Walter Scott, 1892.

[192] J. Perktold, S. Seabold, J. Taylor, and Statsmodels-developers. Fleiss' kappa multi-rater agreement measure. `http://www.statsmodels.org/stable/generated/statsmodels.stats.inter_rater.fleiss_kappa.html`, 2018.

[193] B. Poblete, R. Garcia, M. Mendoza, and A. Jaimes. Do all birds tweet the same?: characterizing Twitter around the World. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1025–1030. ACM, 2011.

[194] V. Rakesh, C. K. Reddy, D. Singh, and M. Ramachandran. Location-specific tweet detection and topic summarization in Twitter. In *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*, pages 1441–1444. IEEE, 2013.

[195] N. M. Razali, Y. B. Wah, et al. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of statistical modeling and analytics*, 2(1):21–33, 2011.

[196] K. Reinecke and A. Bernstein. Improving performance, perceived usability, and aesthetics with culturally adaptive user interfaces. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(2):8, 2011.

[197] K. Reinecke, G. Reif, and A. Bernstein. Cultural user modeling with cumo: An approach to overcome the personalization bootstrapping problem. In *Workshop on Cultural Heritage on the Semantic Web, International Semantic Web Conference*, 2007.

[198] K. Reinecke, S. Schenkel, and A. Bernstein. Modeling a users culture. *Handbook of research on culturally-aware information technology: Perspectives and models, IGI Global, Hershey PA*, 2010.

[199] S. Rendle. Factorization machines. In *2010 IEEE International Conference on Data Mining*, pages 995–1000. IEEE, 2010.

[200] B. Rieder. The refraction chamber: Twitter as sphere and network. *First Monday*, 17(11), 2012.

[201] C. M. Rivers and B. L. Lewis. Ethical research standards in a world of big data [v1; ref status. *F1000Research 2014*, 3(38), 2014.

[202] M. Roblyer, M. McDaniel, M. Webb, J. Herman, and J. V. Witty. Findings on Facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites. *The Internet and Higher Education*, 13(3):134–140, 2010.

[203] S. Roller, M. Speriosu, S. Rallapalli, B. Wing, and J. Baldridge. Supervised text-based geolocation using language models on an adaptive grid. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1500–1510. Association for Computational Linguistics, 2012.

[204] A. Rosen. Tweeting made easier. `https://blog.twitter.com/official/en_us/topics/product/2017/tweetingmadeeasier.html`, 2017.

[205] V. Routamaa, T. M. Hautala, and Y. Tsutzuki. Managing intercultural differences: the relationships between cultures, values and personality. *International Journal of Society Systems Science*, 2(3):269–284, 2010.

[206] M. Rowe. Semanticsvd++: Incorporating semantic taste evolution for predicting ratings. In *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 01*, pages 213–220. IEEE Computer Society, 2014.

[207] A. Said and A. Bellogín. Coherence and inconsistencies in rating behavior: estimating the magic barrier of recommender systems. *User Modeling and User-Adapted Interaction*, 28(2):97–125, 2018.

[208] M. Y. H. Saito and K. Universiry. Multi-language sentiment analysis of SNS across different cities. *The Association for Natural Language Processing*, 2017.

[209] H. Sanchez and S. Kumar. Twitter bullying detection. *ser. NSDI*, 12:15–15, 2011.

[210] J. Schantl, R. Kaiser, C. Wagner, and M. Strohmaier. The utility of social and topical factors in anticipating repliers in Twitter conversations. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 376–385. ACM, 2013.

[211] M. Schedl and D. Hauger. Mining microblogs to infer music artist similarity and cultural listening patterns. In *Proceedings of the 21st international conference companion on World Wide Web*, pages 877–886. ACM, 2012.

[212] P. M. Schwartz. Property, privacy, and personal data. *Harvard Law Review*, pages 2056–2128, 2004.

[213] scikit learn.org. scikit-learn. `http://scikit-learn.org/stable/modules/generated/sklearn.metrics.cohen_kappa_score.html`, 2018.

[214] scipy.org. scipy.stats.levene. `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.levene.html`, 2018.

[215] scipy.org. scipy.stats.normaltest. `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html`, 2018.

[216] scipy.org. scipy.stats.ttest_ind. `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html`, 2018.

[217] F. Sebastiani. Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1):1–47, 2002.

[218] semiocast.com. Brazil becomes 2nd country on Twitter, Japan 3rd Netherlands most active country: Geolocation analysis of Twitter accounts by Semiocast. `https://semiocast.com/publications/2012_01_31_Brazil_becomes_2nd_country_on_Twitter_superseds_Japan`, 2012.

[219] R. B. Severtson, W. A. Rohm, and G. Ericson. Feature engineering in data science. `https://azure.microsoft.com/en-us/documentation/articles/machine-learning-feature-selection-and-engineering`, 2017.

[220] G. Shani and A. Gunawardana. Evaluating recommendation systems. In *Recommender systems handbook*, pages 257–297. Springer, 2011.

[221] A. Shontell. What it's like to sell your startup for $120 million before it's even launched: Meet Twitter's new prized possession, Periscope. `http://www.businessinsider.com/what-is-periscope-and-why-twitter-bought-it-2015-3`, 2015.

[222] N. Shuyo. Language detection library for Java, 2010.

[223] M. Sigala and O. Sakellaridis. Web users'cultural profiles and e-service quality: Internationalization implications for tourism web sites. *Information Technology & Tourism*, 7(1):13–22, 2004.

[224] A. Signorini, A. M. Segre, and P. M. Polgreen. The use of Twitter to track levels of disease activity and public concern in the US during the influenza a H1N1 pandemic. *PloS one*, 6(5):e19467, 2011.

[225] T. H. Silva, P. O. de Melo, J. Almeida, M. Musolesi, and A. Loureiro. You are what you eat (and drink): Identifying cultural boundaries by analyzing food & drink habits in Foursquare. *arXiv preprint arXiv:1404.1009*, 2014.

[226] N. Singh, H. Zhao, and X. Hu. Analyzing the cultural content of Web sites: A cross-national comparision of China, India, Japan, and US. *International Marketing Review*, 22(2):129–146, 2005.

[227] A. Smith, L. Dunckley, T. French, S. Minocha, and Y. Chang. A process model for developing usable cross-cultural websites. *Interacting with computers*, 16(1):63–91, 2004.

[228] P. B. Smith, S. Dugan, and F. Trompenaars. National culture and the values of organizational employees a dimensional analysis across 43 nations. *Journal of cross-cultural psychology*, 27(2):231–264, 1996.

[229] A. Soliman, K. Soltani, J. Yin, A. Padmanabhan, and S. Wang. Social sensing of urban land use based on analysis of Twitter users' mobility patterns. *PLos One*, 12(7):e0181657, 2017.

[230] T. Spangler. Time inc. buys Myspace parent company Viant: Time Inc. expects Viant to contribute $100 million in ad revenue for 2016. `https://variety.com/2016/digital/news/time-inc-myspace-viant-1201703860/`, 2016.

[231] Statista. Most popular social networks worldwide as of april 2018, ranked by number of active users (in millions). `https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/`, 2018.

[232] statisticssolutions.com. Paired sample t-test. `https://www.statisticssolutions.com/manova-analysis-paired-sample-t-test/`, 2018.

[233] B. Stone. Tweet preservation. `https://blog.twitter.com/2010/tweet-preservation`, 2010.

[234] L. Strachan, J. Anderson, M. Sneesby, and M. Evans. Pragmatic user modelling in a commercial software system. *COURSES AND LECTURES-INTERNATIONAL CENTRE FOR MECHANICAL SCIENCES*, pages 189–200, 1997.

[235] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009:4, 2009.

[236] T. Swinke. A unique, culture-aware, personalized learning environment. In *Interactive Collaborative Learning (ICL), 2012 15th International Conference on*, pages 1–7. IEEE, 2012.

[237] Y. Takhteyev, A. Gruzd, and B. Wellman. Geography of Twitter networks. *Social networks*, 34(1):73–81, 2012.

[238] L. Terveen and W. Hill. Beyond recommender systems: Helping people help each other. *HCI in the New Millennium*, 1:487–509, 2001.

[239] The Library of Congress. Twitter donates entire tweet archive to Library of Congress. `http://www.loc.gov/today/pr/2010/10-081.html`, 2010.

[240] H. Thode. Testing for normality New York, 2002.

[241] F. Trompenaars and C. Hampden-Turner. *Riding the waves of culture: Understanding diversity in global business*. Nicholas Brealey Publishing, 2011.

[242] Twitter. API reference index. `https://developer.twitter.com/en/docs/api-reference-index`, 2014.

[243] Twitter. Connecting to a streaming endpoint. `https://dev.twitter.com/streaming/overview/connecting`, 2014.

[244] Twitter. Rules of the road. `https://dev.twitter.com/overview/terms/rules-of-the-road`, 2014.

[245] Twitter. Terms of service. `https://twitter.com/tos`, 2014.

[246] Twitter. Twitter API limits. `https://support.twitter.com/groups/56-policies-violations/topics/237-guidelines/articles/160385-twitter-api-limits`, 2014.

[247] Twitter. Twitter libraries. `https://dev.twitter.com/overview/api/twitter-libraries`, 2014.

[248] Twitter. Twitter privacy policy. `https://twitter.com/privacy`, 2014.

[249] Twitter. Your privacy controls for tailored ads. `https://support.twitter.com/articles/20170405`, 2014.

[250] Twitter. Public streams. `https://dev.twitter.com/streaming/public`, 2015.

[251] M. Usuf, C. Adams, and K. Dingley. A novel framework of e-participation for education sector. In *Proceedings of the 14th European Conference on E-Government (ECEG)*, pages 363–372, 2014.

[252] O. Van Laere, J. Quinn, S. Schockaert, and B. Dhoedt. Spatially aware term selection for geotagging. *Knowledge and Data Engineering, IEEE Transactions on*, 26(1):221–234, 2014.

[253] M. Van Setten, S. Pokraev, and J. Koolwaaij. Context-aware recommendations in the mobile tourist application COMPASS. In *Adaptive hypermedia and adaptive web-based systems*, pages 235–244. Springer, 2004.

[254] A. Vasalou, A. Joinson, and D. Courvoisier. Cultural differences, experience with social networks and the nature of "true commitment" in Facebook. *International journal of human-computer studies*, 68(10):719–728, 2010.

[255] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen. Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1079–1088. ACM, 2010.

[256] A. Virta. Adoption of Twitter in Finland: the young Finnish speakers. `https://www.theseus.fi/bitstream/handle/10024/59131/Anniina_Virta_Thesis.pdf`, 2013.

[257] E. Vitkauskaitė. Cultural adaptation issues in social networking sites. *Economics and Management*, 16:1348–1355, 2011.

[258] W. Wahlster and A. Kobsa. *User models in dialog systems.* Citeseer, 1988.

[259] webfoundation.org. Three challenges for the Web, according to its inventor. `http://webfoundation.org/2017/03/web-turns-28-letter/`.

[260] B. L. Welch. The generalization of student's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35, 1947.

[261] J. Q. Whitman. The two western cultures of privacy: Dignity versus liberty. *Yale Law Journal*, pages 1151–1221, 2004.

[262] B. E. Wiggins. *The impact of cultural dimensions and the coherence principle of multimedia instruction on the achievement of educational objectives within an online learning environment*. PhD thesis, Indiana University of Pennsylvania, 2011.

[263] wikipedia.org. ARPANET. `https://en.wikipedia.org/wiki/ARPANET`, 2018.

[264] wikipedia.org. Clixtr. `https://en.wikipedia.org/wiki/Clixtr`, 2018.

[265] wikipedia.org. Collaborative filtering. `https://en.wikipedia.org/wiki/Collaborative_filtering`, 2018.

[266] wikipedia.org. Discounted cumulative gain. `https://en.wikipedia.org/wiki/Discounted_cumulative_gain`, 2018.

[267] wikipedia.org. Foursquare. `https://en.wikipedia.org/wiki/Foursquare`, 2018.

[268] wikipedia.org. Snapchat. `https://en.wikipedia.org/wiki/Snapchat`, 2018.

[269] wikipedia.org. Student's t-test. `https://en.wikipedia.org/wiki/Student%27s_t-test`, 2018.

[270] wikipedia.org. Timeline of social media. `https://en.wikipedia.org/wiki/Timeline_of_social_media`, 2018.

[271] wikipedia.org. Twitter. `https://en.wikipedia.org/wiki/Twitter`, 2018.

[272] D. Wilkinson and M. Thelwall. Trending Twitter topics in English: An international comparison. *Journal of the American Society for Information Science and Technology*, 2012.

[273] D. Williamson. Forward from a critique of Hofstede's model of national culture. *Human Relations*, 55(11):1373–1395, 2002.

[274] T. Wolf, J. Chaumond, and C. Delangue. Meta-learning a dynamical language model. *arXiv preprint arXiv:1803.10631*, 2018.

[275] P. Xie, Y. Pei, Y. Xie, and E. Xing. Mining user interests from personal photos. *AI Magazine*, 2015.

[276] B. W. Yap and C. H. Sim. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12):2141–2155, 2011.

[277] L. Zhang and W. Zhang. An information extraction attack against on-line social networks. In *Social Informatics (SocialInformatics), 2012 International Conference on*, pages 49–55. IEEE, 2012.

[278] W. Zhang and J. Gelernter. Geocoding location expressions in Twitter messages: A preference learning method. *Journal of Spatial Information Science*, 2014.

[279] W. Zhang, T. J. Johnson, T. Seltzer, and S. L. Bichard. The revolution will be networked the influence of social networking sites on political attitudes and behavior. *Social Science Computer Review*, 28(1):75–92, 2010.

[280] Z.-K. Zhang, C. Liu, Y.-C. Zhang, and T. Zhou. Solving the cold-start problem in recommender systems with social tags. *EPL (Europhysics Letters)*, 92(2):28002, 2010.

[281] X. Zhao. When to use Scott's π or Krippendorff's α, if ever? In *Proceedings of the annual conference of Association for Education in Journalism and Mass Communication*. Association for Education in Journalism and Mass Communication, 2011.

[282] X. Zheng, J. Han, and A. Sun. A survey of location prediction on Twitter. *IEEE Transactions on Knowledge and Data Engineering*, 2018.

[283] Y. Zheng and Y. Yano. A framework of context-awareness support for peer recommendation in the e-learning context. *British Journal of Educational Technology*, 38(2):197–210, 2007.

[284] K. Zhou, S.-H. Yang, and H. Zha. Functional matrix factorizations for cold-start recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 315–324. ACM, 2011.